

# Life after SuperBabe

In the 30 years since the birth of the world's first 'test tube' baby, *in vitro* fertilization has become commonplace. The next three decades could bring equally transformative technologies.

She arrives in the delivery room of a British hospital just before midnight, weighing 3.4 kilograms — a routine birth for a baby who is anything but. Her parents keep a copy of the newspaper to mark her birthday: 25 July 2038.

She is just what they dreamed of, of course, because they did everything medically possible to make sure of it. They had her genome sequenced by plucking off a cell or two when she was an embryo, just as they did for the cluster of other embryos produced by the *in vitro* fertilization (IVF) process. They chose her when the Baby's First Four Letters™ analysis at the clinic said that this particular embryo had the best odds of growing up to be thin, happy and cancer-free...

If this 30-year-hence scenario seems entirely plausible, it is because of what happened 30 years ago this month, when the first baby created by IVF was born on 25 July 1978. The papers called her SuperBabe. Her parents called her Louise Brown. Since then, what once seemed incredible and controversial has become commonplace. Some 4 million babies have already been born via IVF. So in this issue, *Nature* asks experts in reproductive medicine to speculate on what the next three decades might hold (see page 260). Some of the techniques promise to be equally transformative, if they come to pass.

Consider, for example, what would happen if researchers learn to grow artificial sperm and eggs from other body cells (see *Nature* 452, 913; 2008). They would have abundant raw material for IVF, and could potentially bring about an end to infertility altogether. As that scenario would also lead to a bountiful supply of embryos, genetic screening could become a necessity — and the door would open wider to allow genetic enhancement and modification of germ cells and embryos.

Already, modern societies are entering an era of personalized genetics, in which anyone can pay for a read-out of known risk genes — or, soon, a complete personal genome sequence. These technologies will make their way into the fertility clinic. True, with thousands of genetic risk variants contributing to multiple different conditions, no embryo will have the perfect genetic future. But these techniques could allow parents to create a top-five wish-list of the characteristics they most

want for their child — avoiding, for example, the Parkinson's disease that plagues the family — and choose the embryo most likely to meet those criteria. Or the parents may focus on non-health-related aspects such as intelligence and ambition; the ethical debate about genetic selection is likely to intensify over the next few years, as it should.

Meanwhile, safety concerns about IVF have still not evaporated, even after 30 years. Although it is unlikely that IVF does any major harm, more subtle problems may become apparent only when very large numbers of children are followed into middle age or beyond. Yet few such studies have been initiated. There are almost no large registries tracking children born via IVF, and even less information on children subjected to more recent techniques such as preimplantation genetic diagnosis. Such long-term studies are expensive and difficult because the privacy of parents and children must be maintained, and many will choose not to participate. Nonetheless, such registries should be a priority — even more so as the next generation of assisted reproduction techniques comes online. Yes, prospective parents may have to accept risks — but they should at least know what those risks are.

Also not resolved in the past 30 years is how to ensure that the appropriate safety and ethical requirements are satisfied. One model is Britain's widely admired Human Fertilisation and Embryology Authority, which has the legislative backing to set rules and enforce them (see page 280). In the absence of such regulation, as in the United States, the onus is on doctors to prove that they are committed to transparency, safety and the best outcome for both prospective parents and their children.

What is certain is that our future newborn on her birthday will be oblivious to these debates and to the method of her creation. Her existence will demonstrate that nothing is sacred in human biology — and researchers should ensure that nothing is diminished about human reproduction by starting it in the lab. ■

**"Already, modern societies are entering an era of personalized genetics."**

## Templeton's legacy

The Templeton Foundation's exploration of science and faith merits tolerance, not outright rejection.

When a wealthy individual seeks to leave a legacy through scientific philanthropy, researchers usually greet such generosity enthusiastically. But the death of investment mogul John Templeton marks an unusual, and notable, exception. At the time of his passing last week, Templeton had poured some US\$1.5 billion into the John Templeton Foundation, which funds

research at the intersection of science and spirituality. Critics have maintained that the foundation needlessly conflates science and faith, with some calling for an outright boycott of Templeton funding.

Templeton was a deeply spiritual, albeit unorthodox, individual (see page 290). He lived a life firmly rooted in the Christian traditions of modesty and charity. Yet he was also a great admirer of science, the undogmatic practice of which he believed led to intellectual humility. His love of science and his God led him to form his foundation in 1987 on the basis that a mutual dialogue might enrich the understanding of both.

This publication would turn away from religion in seeking explanations for how the world works, and believes that science is

likely to go further in explaining human moral impulses than some religious people will welcome. Thus it shares a degree of suspicion with many in the scientific community at any attempt by religiously driven organizations to fund science. A chief concern is that the influential Templeton Foundation might be seeking to inject religion into the scientific world. And it is easy to understand that concern given the political activism of many American fundamentalists and their efforts to promote ideas such as intelligent design, which posits a divine hand in evolution. The foundation's most vigorous critics accuse it of attempting to lace science with spiritualism.

That claim is somewhat ironic, as Templeton himself seemed to have just the opposite in mind. He believed institutional religion to be antiquated, and hoped a dialogue with researchers might bring about advances in theological thinking. The foundation's substantial funding of science and religion departments around the world is directed towards those ends. Theologians have also used foundation money to develop and promote arguments that reconcile some of the apparent contradictions between science and religion. For those many scientists with a faith, promoting the compatibility of science with faith is a prudent and even necessary goal. Strict atheists may deplore such activities, but they can happily ignore them too.

The foundation's scientific agenda addresses 'big questions', which has sometimes resulted in work that many researchers regard as

scientifically marginal. One field popular with the foundation is positive psychology, which seeks to gauge the effects of positive thinking on patients, and which critics argue has yielded little. Also heavily supported are cosmological studies into the existence of multiple universes — a notion frequently criticized for lying at the edge of falsifiability. The concern is that such research has been unduly elevated by the foundation's backing. But whatever one thinks of positive psychology and the like, the foundation's support has not taken anything away from conventional funding. And in the field of cosmology at least, it has arguably yielded some new and interesting ideas.

**"Templeton believed institutional religion to be antiquated, and hoped a dialogue with researchers might bring about advances in theological thinking."**

The foundation's management now falls chiefly to Templeton's son, John M. Templeton Jr, whose Christian beliefs are reportedly much more conventional than his father's. A critical scrutiny of the foundation's scientific influence continues to be warranted, and no scientific organization should accept sums of money so large that its mission could be perceived as being swayed by religious or spiritual considerations. But critics' total opposition to the Templeton Foundation's unusual mix of science and spirituality is unwarranted. ■

## An uneasy peace

Britain's 'big science' funding agency is now in a position to regain much-needed credibility.

Last week, an official opened a meeting between scientists and the UK Science and Technology Facilities Council (STFC) by asking that those present leave their weapons at the front desk.

The joke, which met with anaemic laughter, shows how bad things have been between the council, whose responsibilities include high-energy physics and astronomy, and the scientists it serves. In December, the council announced that it had an £80-million (US\$160-million) spending shortfall in its latest budget, which runs until 2011. Council officials laid out preliminary plans to withdraw from such key projects as the International Linear Collider, a next-generation particle accelerator, and the Gemini Observatory, a pair of 8-metre telescopes located in Hawaii and Chile. Many were furious over the cuts, which came with no consultation.

Gallows humour aside, last week's meeting showed that the STFC has gone some considerable way towards repairing its relationship with the community. Resentment remains, especially towards Keith Mason, the council's sometimes truculent chief executive. But by and large, the researchers who depend on the STFC to back their work seem ready to accept a programme that includes some cuts. This transformation is thanks to the rapid formation of ten specialist advisory committees to help inform the final version of the STFC's budget.

Although the plan looks similar to the original package, important concessions have been made and priorities shifted in a way that has

ameliorated the community's initial rage. The final plan sets aside around £1 million for 'advanced detector work', similar to that being done in preparation for the linear collider. It also continues participation in the Gemini telescopes, although it will seek to sell half of Britain's observing time in the project. The plan also promises support to projects in other fields, such as nuclear and neutrino physics.

The truce between community and council comes just in time. Already the UK government is gearing up for its next budget review, and the STFC and its constituent physicists must be able to work in concert if they are to win a bigger slice of the cake in the next round. They must speak with a single voice to policy-makers about the broader value of their work, and they must be coherent about the consequences of lower funding levels.

Coming up with a consistent message will not be easy. The STFC supports many disciplines. Yet at last week's meeting there was a sense of common purpose. The message from both the crowd and the STFC was that their work and especially the people who do it provide an intellectual foundation on which the knowledge economy is built. That message should resonate reasonably well with the Treasury, which is seeking economic returns on its investment in science, and will ring more true to scientists than promises of spin-off technologies and business-government partnerships.

It is up to both sides to develop last week's germ of an idea into a full-blown campaign. The STFC can work with the community to communicate effectively to policy-makers, while researchers, through the newly formed advisory committees, must tell the council how their work fits with the broader goals of the STFC. A dialogue of this sort, sorely absent this past eight months, is essential if the funding shortfalls seen last winter are not to be repeated in the next spending cycle. ■

# RESEARCH HIGHLIGHTS

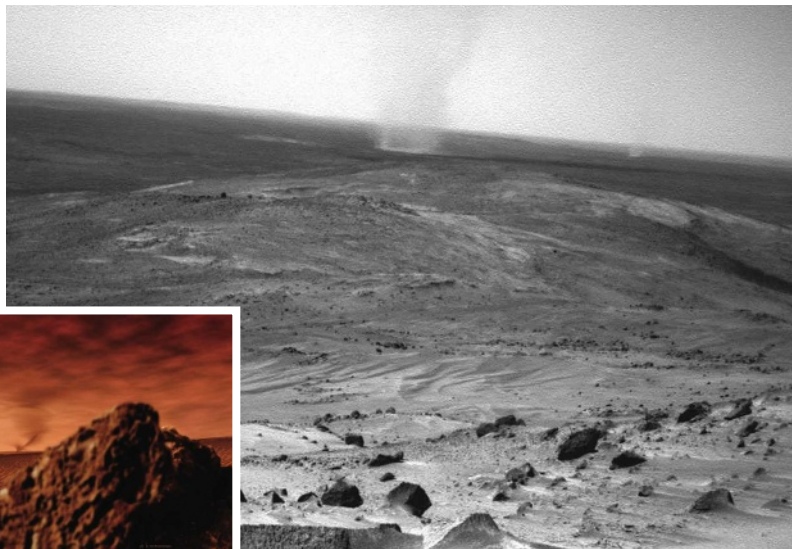
## Martian devilry

*J. Geophys. Res.* doi:10.1029/2007JE002966 (2008)

Dust devils — dancing, twisting vortices that suck dirt high into the air — seem to contribute a surprisingly large puff of dust to the martian atmosphere.

Patrick Whelley and Ronald Greeley of Arizona State University in Tempe counted almost 55,000 tracks left by dust devils over the course of three years by scrutinizing images from the Mars Orbital Camera. The devils mainly clumped in narrow, seasonally dependent bands at around 60° North and South.

Altogether, dust devils contribute roughly half as much dust to the atmospheric total as do bigger dust storms — information that could improve models that help spacecraft to navigate their way to the red planet's surface. The images show an artist's impression (inset) and a dust devil photographed by NASA's *Spirit* rover.



D. VAN RAVENSWAAY/SPL/NASA/JPL

## MATERIALS SCIENCE

### All white

*J. Am. Chem. Soc.* **130**, 8114–8115 (2008)

Light-emitting diodes (LEDs) that emit white light could offer low-energy indoor lighting. But they are hard to make. Today's examples rely on red–green–blue mixtures or complex blends of materials, which are cumbersome and costly.

Wooseok Ki and Jing Li of Rutgers University in Piscataway, New Jersey, have discovered a new type of semiconducting material that emits white light when illuminated by a blue LED. It consists of stacked sheets of cadmium sulphide that have amine molecules attached to their top and bottom surfaces.

The brightness is improved by adding a little manganese. Next the researchers will need to see if the white emission can be driven purely by electricity.

Loya and Sakai note that this coral's size-dependent, labile sexuality is remarkably similar to that of a Japanese tree, *Bischofia javanica*.

## GENETICS

### Autistic details

*Science* **321**, 218–223 (2008)

A pedigree analysis of 104 families has revealed associations between several genes and disorders in the spectrum that includes autism and Asperger syndrome.

Christopher Walsh of Harvard Medical School in Boston, Massachusetts, and his colleagues collected data from Middle Eastern families with affected members. Eighty-eight families in their sample had at least one child-bearing marriage between first cousins, which makes mapping genetic conditions simpler.

The analysis revealed several genomic regions that are linked to autism, including

five large deletions. Several genes within these regions are thought to have a role in learning and seem to be important for stages of brain development during which autism symptoms usually develop.

## VIROLOGY

### Collective calm

*Biophys. J.* doi:10.1529/biophysj.108.133694 (2008)

Whether viruses kill a host or lie dormant may be decided 'by committee', according to a model designed by Joshua Weitz at the Georgia Institute of Technology in Atlanta and his collaborators.

Viruses that infect bacteria can 'choose' to burst out of the host cell or to sit tight in a latent state. Previous research with the virus bacteriophage  $\lambda$  has shown that if two or more virus particles inhabit a single cell they are more likely to defer killing their host than is a lone virus.

The more virus particles there are in a cell, the greater the overall level of viral gene expression. Weitz's model proposes that even slight increases in the viral messenger RNA beyond the amount that can be produced from a single virus can have a dramatic effect on gene networks that control cell fate.

## ZOOLOGY

### Quick-change artists

*Proc. R. Soc. B* doi:10.1098/rspb.2008.0675 (2008)

Although many animals can switch from one sex to the other, repeated flip-flopping was thought to be the preserve of one species of fish. But some corals do it too, reveal Yossi Loya at Tel Aviv University and Kazuhiko Sakai at the University of the Ryukyus in Japan.

They collected wild mushroom coral, *Ctenactis echinata* (pictured right), and monitored it in a laboratory over many months. Large coral heads remained female and small ones remained male, but medium-sized corals changed sex bidirectionally and often.



## GEOSCIENCE

### Carbon sinks

*Geophys. Res. Lett.* doi:10.1029/2008GL034271 (2008)

The contents of a sediment trap suspended more than 3,000 metres below the sea surface may force a reassessment of the Arctic Ocean's carbon cycle.

Jeomshik Hwang and his colleagues from



Woods Hole Oceanographic Institution in Massachusetts report that organic carbon caught in their sediment trap in the Canada Basin is “strikingly old”, with an approximate age of 2,000 years. If the carbon in the trap came from particulate matter descending from the surface, as oceanographers would expect, its age should be indistinguishable from that of surface carbon.

The findings imply that organic carbon at a depth of 3,000 metres largely originates from the basin's surrounding margins. This sets the Canada Basin apart from other ocean basins studied and means that models generally used to describe ocean carbon cycling do not apply in the Arctic.

## BIOCHEMISTRY

### Cook the catalyst

*J. Am. Chem. Soc.* doi:10.1021/ja802404g (2008)  
Microwaving enzymes that work best in hot environments can boost their activity at near-room temperature, find Alexander Deiters and his co-workers at North Carolina State University in Raleigh. They exposed several enzymes from hyperthermophilic organisms that function best at 90–110 °C to microwave radiation that warmed the enzyme–substrate mixture to around 40 °C.

Normally, the enzymes would have done little at 40 °C, but the microwaves multiplied their industriousness at this temperature, in one case more than fourfold. Deiters and his team attribute the effect to a loosening up of the enzymes' molecular structures, caused by interactions of the molecules and the microwaves' oscillating electric field.

## MOLECULAR IMAGING

### A gentler touch

*Phys. Rev. Lett.* **101**, 013001 (2008)  
In order to watch chemical reactions as they happen, researchers need the constituent molecules lined up just right. Lasers can help to achieve this, but usually disturb the reaction process because they excite the molecules.

Hirofumi Sakai and his colleagues at the University of Tokyo now demonstrate a way to avoid this excitation with carbonyl sulphide molecules. First, they placed the molecules in a weak electrostatic field. Then they zapped the carbonyl sulphide with a nanosecond laser pulse shaped like a cresting wave. The pulse's shape nudged the molecules into alignment, where they remained after it was turned off.

The team believes that this technique could be adapted for studies of reaction dynamics and molecular imaging.

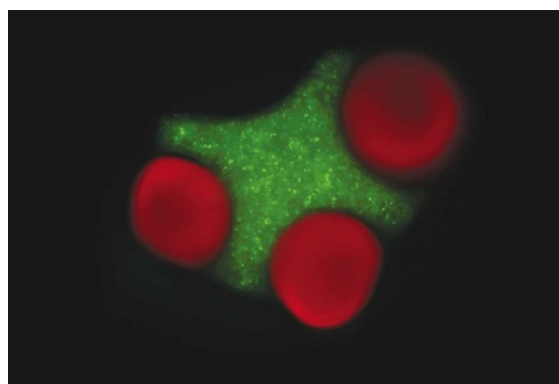
## TISSUE ENGINEERING

### To rig with oil

*Proc. Natl Acad. Sci. USA* **105**, 9522–9527 (2008)  
Three-dimensional artificial tissue structures can assemble themselves when shaken with a little oil, report Ali Khademhosseini at Harvard Medical School in Cambridge, Massachusetts, and his co-workers.

The researchers made use of the tendency of water and oil to repel one other to force water-loving microgels — polymers packed with cells — to form straight lines and other shapes (such as the cross attached to three rods, pictured below).

The hydrophilic microgels minimized their surfaces' contact with the oil by stacking closely together. The structures were later fixed in place by a cross-linking reaction.



A. KHADEMHOSEINI

## EVOLUTION

### Sea skeletons

*Nature Geosci.* doi:10.1038/ngeo251 (2008)  
Corals, sponges, bivalves and other calcifying marine creatures secrete their calcium carbonate shells and skeletons in the form of aragonite, calcite or both. The evolutionary successes of aragonitic and calcitic genera over the past 500 million years were influenced much more strongly by mass extinctions than by the chemical composition of the seas, researchers have found.

This is surprising because the oceanic ratio of magnesium to calcium has fluctuated throughout this period; a high ratio facilitates aragonite precipitation, a low one promotes calcite precipitation. But Wolfgang Kiessling of Humboldt University in Berlin, Germany, and his colleagues show that this cycling does not match the relative abundances of aragonite- and calcite-secretors in the fossil record. The mineralization changes correlate with mass extinctions, which seem to alter the fortunes of the two categories at random.

## JOURNAL CLUB

**Seth Putterman**  
University of California, Los Angeles

### A physicist links magnetism, force and fatigue.

If a metal bar is repeatedly stretched and released it becomes fatigued and, eventually, ruptures. The latter can occur suddenly and unexpectedly: sometimes materials scientists can find no obvious thermodynamic hint that a steel rod is about to break. I am interested in fatigue because it parallels other phenomena that concentrate energy density, such as triboluminescence, whereby diffuse stress makes a crystal glow.

In both triboluminescence and fatigue, applied forces cause molecular rearrangements. But fatigue also involves nanometre-sized defects that accumulate during the useful life of a piece of metal and organize themselves into a soft spot. Recently, Sidney Guralnick and his colleagues at the Illinois Institute of Technology in Chicago measured how much work is needed to complete each 'stretch and release' cycle in rods of AISI 1018 steel, a common low-carbon steel that is used in vehicle parts such as gears (S. A. Guralnick *et al. J. Phys. D Appl. Phys.* **41**, 115006; 2008). This allowed them to follow changes in the material's response to force as it fatigued.

A shift occurred at merely 12.3% of the time to rupture. What is happening inside the steel at this point is mysterious, but the number holds true even when the useful life of identically manufactured rods varies by a factor of 200.

Further clues will no doubt come from steel's piezomagnetism — the fact that its magnetism varies with the degree of stretch it experiences. This relationship is complex: even when the metal is so slightly strained that it goes back to its original shape on release, its magnetic field does not return to the pre-stretched state. One day investigations into this property may uncover the organizing principle of the nanometre-sized defects that underlie metal rupture.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

# Russia delays Lake Vostok drill

## ST PETERSBURG

Russia is postponing its controversial plans to drill into Antarctica's sub-glacial Lake Vostok. Russian scientists now hope to start probing the pristine environment almost 4,000 metres below the East Antarctic ice sheet in the 2009–10 drilling season.

"For technical and legal reasons, penetration is not yet possible this coming season," Valery Lukin, director of the Russian Antarctic Expedition, told *Nature* last week at a meeting in St Petersburg of the Scientific Committee on Antarctic Research (SCAR).

Drilling, which got stuck last year in the accretion ice — lake water frozen onto the bottom of the ice sheet — some 80 metres above the lake surface, will be resumed in November, he says. But before attempting the final entry into the lake, Russia will submit a finalized environmental evaluation to the 46 members of the Antarctic Treaty at their April 2009 meeting in Baltimore, Maryland. The evaluation contains responses to concerns raised by scientists who are fearful that the drilling will contaminate the uniquely unspoiled environment.

Stopping the drill 30 metres or so above the lake surface, as SCAR recommended more than ten years ago, is not seen as an option: the Vostok project is central to the whole Russian Antarctic programme, and is a matter of national importance similar to the race to the Moon in the 1960s, says Lukin. "We will definitely drill into the lake in 2010, with Russian-developed



Lake Vostok lies thousands of metres below Vostok Station, the coldest recorded place on Earth.

RIA NOVOSTI/SPL

drilling technology," he says. "But most importantly, we'll do so in full compliance with all the rules of national and international law."

The legitimacy of the drilling is undisputed, but concerns remain about possible biological and chemical contamination of the lake. "The Russians have met their obligations under the Antarctic Treaty," says Chuck Kennicutt, the US delegate to, and a vice-president of, SCAR. "Ideally, they would first test the drilling technology at a less sensitive location. But there is no way of preventing them from going ahead."

Lake Vostok lies 1,200 kilometres inland,

beneath one of Russia's Antarctic stations, in one of the coldest parts of the continent. It is the largest and deepest of the 150 or so subglacial freshwater lakes that have been discovered by radar imagery beneath the Antarctic ice sheet. Similar in size to Lake Ontario, it contains more than 5,000 cubic kilometres of up to 1-million-year-old water. This water is super-saturated with oxygen and is liquid below its normal freezing point, because of the high pressure beneath the ice. Scientists believe that the sealed lake, if it is able to support microbial life, might offer a unique window into ancient DNA.

# Pakistan finds two radioactive containers

When Pakistan ran an advertisement campaign last year telling citizens what to do if they stumbled upon radioactive material, it raised eyebrows in the West. Now Pakistani authorities have discovered two radioactive containers of unknown origin, bringing into further doubt their ability to keep track of nuclear material.

The sealed containers were dug up from under a workshop by staff at the state-run Oil and Gas Development Corporation in Karachi. The Pakistan Nuclear Regulatory Authority (PNRA) categorized them as 'orphan

sources' last week, as the oil company have no record of where they came from.

It's not unheard of for misplaced radioactive sources, for example those originally used in industrial processes, to pop up from time to time. These orphan sources usually cause little consternation in the international community; but in this case, there are added worries over their origin and the security of Pakistan's nuclear programme. "Pakistan doesn't have a good record for keeping track of where radioactive materials are going," says Peter Crail, a research analyst at the Arms Control Association,

a non-profit organization in Washington DC. "The 1990s saw a lot of illicit trafficking of radioactive materials between this region of Asia and eastern Europe."

While there has been a clampdown on trafficking since 2000, such material could pose a threat if it were to fall into the wrong hands and be used for weapons, says Crail. Last year's ad campaign launched by the Pakistani authorities to help the public recognize stray radioactive material does little to inspire confidence, he adds. "It suggests that their ability to make sure they know where these sources are is

not up to par."

"The worry is the lack of security of the Pakistani nuclear programme," says Jeffrey Lewis, director of nuclear strategy and non-proliferation at the New America Foundation in Washington DC. It is important to determine what the containers hold and where they came from, he adds.

The material has not been identified, but the PNRA states that it is a neutron source with low radioactivity and that such materials are commonly used in oil exploration. The site at which they were unearthed was used by a Soviet oil company until the late

**GOT A NEWS TIP?**

Send any article ideas for Nature's News section to [newstips@nature.com](mailto:newstips@nature.com)

K. CAMPBELL/GETTY IMAGES

But accidental contamination could irretrievably spoil this potential ecosystem. Some contamination of the upper water layer is unavoidable when the lake is entered, says Sergey Bulat, a molecular biologist at the Petersburg Nuclear Physics Institute, where Vostok ice cores are being analysed, and where water probes will first arrive. "I don't believe the upper layer, which is basically glacier water, contains any life anyway," he says.

To attempt contamination-free sampling of the middle and bottom layers of the lake, Russian scientists plan to release a small robotic vehicle through the transportation module in the drilling device. The Russian Academy of Sciences is expected to approve funding of around US\$5 million for the 'hydrobot' programme.

Speculation about life in the virtually unexplored lake caught the attention of scientists and the public early on. But the discovery in the past few years of numerous other lakes beneath the Antarctic ice sheet, and of streams of subglacial water flowing rapidly between reservoirs, has deprived Lake Vostok of its unique status. Several proposals for drilling into moving water beneath the west Antarctic ice sheet are currently under review in the United States, for example.

"I'm not as stressed any more about Lake Vostok," says geologist Robin Bell, of Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York. "It's still special, but it is reassuring to know that there are other subglacial environments we can preserve by not probing them at all." ■  
Quirin Schiermeier

1960s. "In the 1960s, there was no nuclear regulatory authority in the country," PNRA spokesman Zaheer Baig told *Nature*. "We think that when it was winding up its business, the oil company buried the source, covered in heavy shielding, for the public's safety." The detected radiation level from the sources is only "slightly higher than background", he adds.

If this is the case, the radiation levels from the containers will probably have been relatively harmless, says Matthew Bunn, a non-proliferation expert at Harvard University. "Sources of modest size are used in the oil industry all over the world, and it is not unusual to find that someone tries to save money by disposing of them improperly," he says. "It is yet another indicator, however, of the need for better cradle-to-grave tracking and control of radiological sources all over the world." ■  
Zeeya Merali

## Autism study panned by critics

The leading US government funder of autism research is drawing fire over its proposal to run a randomized clinical trial of a treatment widely viewed by experts to be useless and potentially harmful, but that is broadly used for autism.

Chelation therapy, in which agents such as dimercaptosuccinic acid (DMSA) are used to bind metal ions in the blood so that they can be excreted easily, is an approved treatment for heavy-metal poisoning. Parents are using such therapy on children with autism because of their belief — which has been scientifically discredited — that mercury from vaccinations caused their children's condition.

In May, investigators at the US National Institute of Mental Health (NIMH) in Bethesda, Maryland, won approval from the US Food and Drug Administration to use DMSA in a trial of children with autism who are aged four to ten years and have detectable, but not toxic, levels of mercury or lead in their blood. The trial, 'Mercury Chelation to Treat Autism', is now under ethics review and has not enrolled any patients.

Critics say the trial will put children at risk for what is certain to be no medical gain. The American Academy of Pediatrics has concluded that there is no justification for giving children DMSA in the absence of very high levels of heavy-metal exposure, notes epidemiologist Ellen Silbergeld of the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. "I don't know why we have to do this experiment again on children."

Louis Cooper, a vaccine expert and former president of the American Academy of Pediatrics, says that despite a growing scientific consensus that mercury in vaccines does not cause autism, many parents still think that it is a problem. He says that a well-designed study of chelation in children with autism would "respond to these parents' deeply held beliefs in the most careful, ethical way". Cooper explains that, with informed consent, such a study could ensure that the only families to be enrolled would be those already determined to try chelation.

Others argue that the study doesn't make scientific sense because autism has not been documented as a symptom of high-dose mercury poisoning and that, even if it were, the damage that mercury does to cells is



R. FARIS/CORBIS

Some parents blame their children's autism on mercury in vaccines.

permanent and not reversible; chelation simply prevents additional harm. They say that the NIMH is bowing to political pressure because a growing number of parents with children who have autism use chelation therapy or want to use it. NIMH director Tom Insel denies pandering to families and says that the idea for the study "came up in the first place because we were getting reports that this was a therapy in broad use and there were very substantial questions about both

its efficacy and its safety". Because chelators bind indiscriminately to metal ions, they can deplete the body of essential metals such as copper, zinc, selenium and calcium. In

Portersville, Pennsylvania, in 2005, a 5-year-old boy with autism died from cardiac arrest after being injected with a chelation agent. (The proposed study uses an oral agent.)

Recruitment for the study, initially proposed in 2006 by Susan Swedo, who heads autism research at the NIMH, was put on hold and additional review undertaken after a report in February 2007 showed that chelating agents could cause cognitive problems in rats (D. Stangle *et al. Environ. Health Perspect.* 115, 201–209; 2007).

Last week Insel said that because the study involves children, carries more than minimal risk and offers no demonstrable benefit to the participants, regulations dictate that it must be referred to a US Department of Health and Human Services panel for ethics approval.

Insel says that the study's in-limbo status shouldn't be read as its demise. "It hasn't been killed by any means. There's certainly a possibility that this project will go forward."

Meredith Wadman



## SPECIAL REPORT

# Making babies: the next 30 years



Louise Brown, the first test-tube baby, was born 30 years ago this month after being conceived outside the body using *in vitro* fertilization (IVF). **Helen Pearson** asks what developments in reproductive medicine could have an equivalent impact in the next three decades.

**Davor Solter, developmental biologist at the Institute of Medical Biology (IMB) in Singapore**

The goals will remain the same in that we'll be trying to give children to those who can't have them and remove children from those who don't want them. I think IVF has gone about as far as it can.

Next I expect that germ cells — sperm and eggs — will be successfully derived from induced pluripotent stem (iPS) cells [cells that have the potential to develop into any of the body's cell types]. It will be possible to make iPS cells from skin cells, to make germ cells from these, and then combine them to make human embryos.

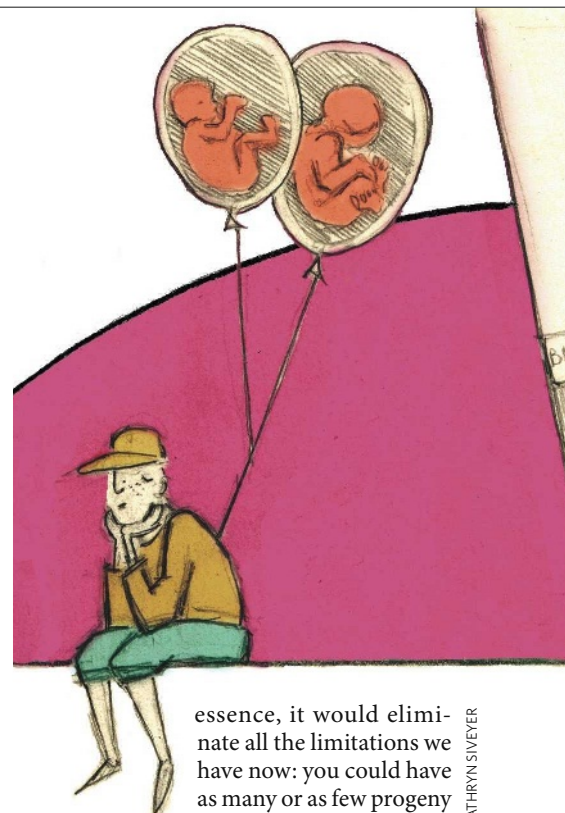
It means every person regardless of age will be able to have children: newborn children could have children and 100-year olds could have children. It could easily happen in the next 30 years.

I have no idea if the technique will be used, but it means you could have millions of

gametes that could be combined at will. Today you can't experiment on human embryos because it's considered morally repugnant — and they are difficult to get. If embryos could be grown in culture like any other cell line, this latter problem would disappear. It would mean you could introduce any kind of genetic modification. The cell lines could be used to correct a mutation or to engineer an improvement, and used to make a mutant embryo for research purposes. They'd become like any other type of cell line. They would become objects and would be used as objects.

I have no idea what kind of moral value or rights we would give to those embryos. We'll probably go through the same agonizing we did with IVF. It could be terrible to begin with, but then it'll become a fact of life. Maybe 20–30 years from now we'll read in newspapers that someone made 20,000 embryos and studied their development, and we'll decide it's OK.

Another thing I predict for this brave new world is the use of artificial placentas. In



essence, it would eliminate all the limitations we have now: you could have as many or as few progeny as you want. I have no idea

how easy it would be. I can visualize a fetus floating freely in fluid and the umbilical cord attached to a machine. But I don't know how much implantation is necessary. The fetus floats freely, but for the blastocyst [early-stage embryo], the placenta might be necessary for normal morphogenesis [organ development, for example].

ILLUSTRATIONS BY: KATHRYN SIVEYER

## Looking back

**Howard Jones, professor emeritus at Eastern Virginia Medical School in Norfolk who in 1981 produced the United States' first IVF baby**

At that time, we doctors could help only around half of the couples that came to us with fertility problems. We mainly used surgical or endocrinological methods.

I first met Bob Edwards in 1964 when he came to Johns Hopkins, where I was an assistant professor, and he asked for some human eggs because in the United Kingdom they were having trouble getting them.

Five years earlier, Min-Chueh Chang had shown that IVF was possible in rabbits. But human sperm had to be 'capacitated'

before it could be used to fertilize the egg, and this seemed to occur in the female genital tract. Bob's idea was to capacitate the sperm. My job was to give Bob parts of the female genital tract such as cervical mucus and bits of uterus lining to see what worked. In retrospect I think *in vitro* fertilization did actually occur at Hopkins in 1965.

Bob went back to England and kept trying. Capacitation was not the problem as it turned out. The key was to get a culture medium agreeable — it had to do with the acidity and oxygen levels. It turns out the human requires a slightly different culture medium from animals.

It took Bob some 15 years of

trying until Louise Brown was born. We said we'd try for three years and then reconsider. We had 41 attempts with no pregnancies at all. Then, in 1981, my wife suggested we use pergonal to stimulate egg production. We did it and the 13th attempt was successful. These drugs are now standard throughout the world and my wife deserves the credit for that. It worked in about 18 months: the baby was called Elizabeth Carr, she was born in December 1981 and I keep in touch with her. There had been about a dozen IVF babies by then.

In the United States there was a hearing to get a 'certificate of need', which is required for all new procedures. It went on from 2

till 8 p.m. and the 'right-to-lifers' brought in people to testify, and we brought in scientists to testify. The main argument they had was that as several embryos were implanted and only one developed, IVF was causing the others to abort. There was lots of publicity and subsidiary hearings before we got our certificate of need.

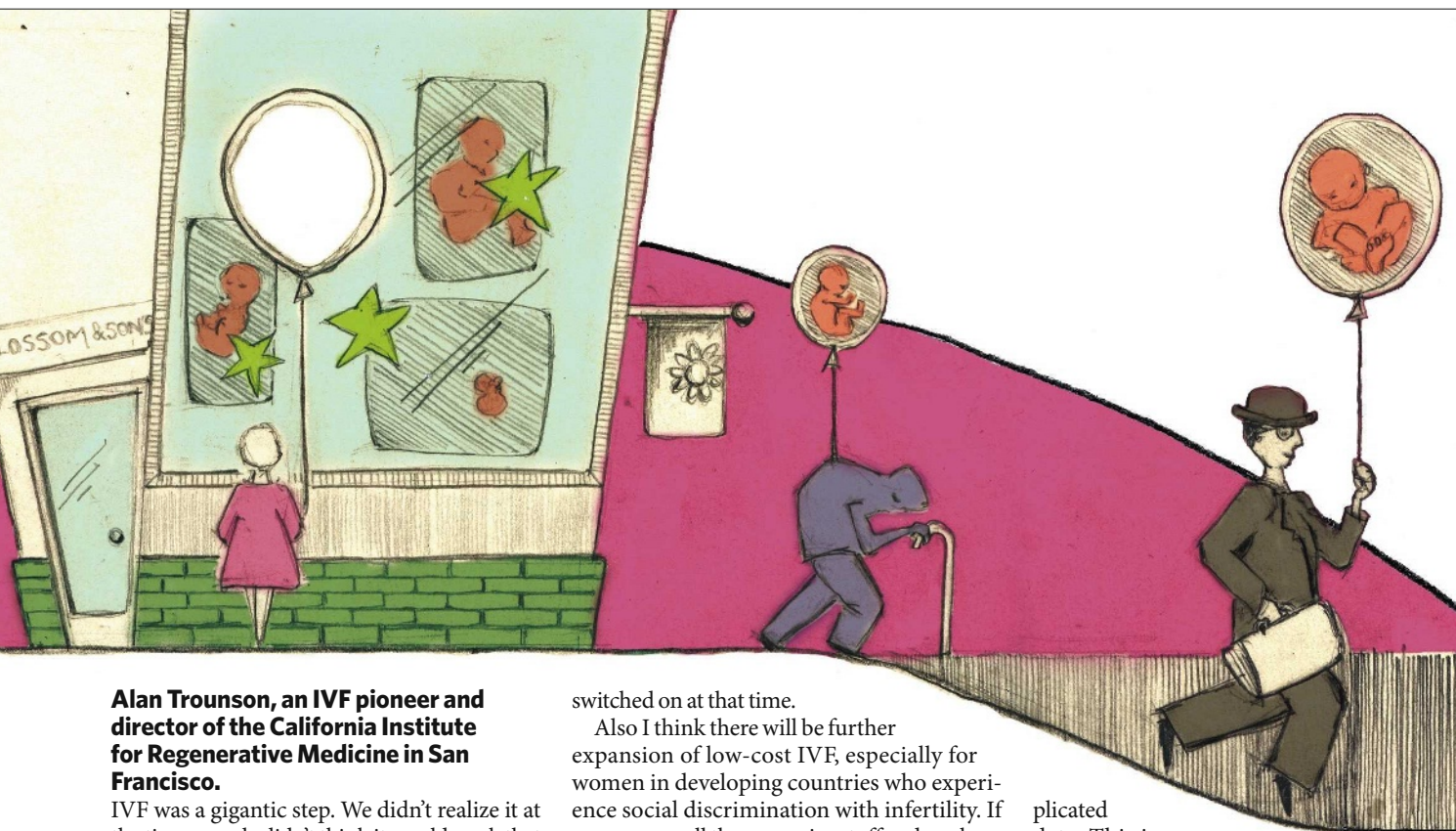
It interests me that society has always wanted to regulate IVF processes — IVF is unique in that regard to all other procedures, such as heart or lung surgery — why is that? What are they trying to regulate? It probably has something to do with the mysteries of reproduction. I think we should move on and treat it like any other procedure. **H.P.**



**HAVE YOUR SAY**

Comment on any of our news stories, online.

[www.nature.com/news](http://www.nature.com/news)



**Alan Trounson, an IVF pioneer and director of the California Institute for Regenerative Medicine in San Francisco.**

IVF was a gigantic step. We didn't realize it at the time; people didn't think it would work that well. We never envisaged that it would expand so dramatically around the world.

Ethics keeps moving. What was once seen as dangerous goes on to be seen as within the confines of acceptable risk. Probably the major development in the field will come from something we've never thought about.

I think it will be possible that we'll be able to extend the fertile period for women by producing germ cells from iPS technology, or by a variant of nuclear transfer, so somatic cells [which make up most of the body's cells] become germ cells and are refreshed genetically. I think this will happen, and it raises a lot of ethical issues about safety and the children. There will be concerns raised over whether the fertile period should be extended beyond its natural point. I think people should be given the choice.

The idea needs first to be established in animal models. We need to look at lifespan because we don't know the impact of reprogramming cells in this way. Another issue is whether we should bank cells from people when they are very young to avoid the accumulation of genetic errors that occur as we get older.

We might see a move towards artificial chromosomes and 'genetic cassettes' that can be inserted at the embryonic stage to correct particular diseases, such as Huntington's. These might be inducible cassettes that can correct for an abnormality that occurs late in life and

switched on at that time.

Also I think there will be further expansion of low-cost IVF, especially for women in developing countries who experience social discrimination with infertility. If you remove all the expensive stuff and use low-cost drugs (such as clomiphene) and remove just one or two eggs, and only transfer one embryo, it can be done for less than US\$100.

**Susannah Baruch, director of reproductive genetics at the Genetics and Public Policy Center at Johns Hopkins University in Washington DC**

Right now preimplantation genetic diagnosis (PGD) is used during IVF to look for single gene disorders in, for example, a family in which a child has already died. I think genetics will continue to play a big part in assisted reproductive technology. The bigger question is: what will people be testing for?

There's speculation that people will have designer babies, but I don't think the data are there to support that. The spectre of people wanting the perfect child is based on a false premise. No single gene predicts blondness or thinness or height or whatever the 'perfect baby' looks like. You might find genetic contributors but there are so many environmental factors too.

More likely is that you'll have a set of embryos and you'll know every single thing about every gene in every embryo. For example, one embryo will have three genes associated with tallness, two for weakness, three for poor vision and some for disease; and the second embryo will have some other set. They're very com-

plicated data. This is not creating a baby from scratch. None of us is a perfect specimen and none of our embryos will be either.

I think you'll end up with a lot of information available to parents and it's less obvious how useful that information will be and how many parents will want it. I think the technology will be there within ten years to be able to know without harming the embryo, but I don't know whether it will become widely used. There are bound to be some parents who want it, and then the dilemma is deciding what you tell your children about the choices you made.

Will people choose IVF to get that genetic option? IVF is expensive and uncomfortable. The old-fashioned way is cheaper and more fun and that won't change in 30 years.

**Alastair Sutcliffe, a paediatrician who studies the health of children after IVF at University College London**

Litigation is a big driver in medicine. With the increasing availability of IVF, there will be more emphasis on safety. Not enough is known about the long-term health of the Louise Browns of this world — if there is a problem, it will be unexpected.

The concern is over whether the culture media used to nourish the cells and embryo before implantation has epigenetic effects [changes to the way a person's genes are expressed]. For

example, there is some evidence that Beckwith Wiedemann syndrome is more common in children conceived through IVF, but whether this is a true effect is unclear. I always say it's rare — but if your child's got it ...

And the fertility of those IVF children as they grow up is of interest. What are the subtle effects? I want to look at epigenetic markers in these people. I'm one of the few people who can do this because of the cohort of children that I have. I think it's a quantum leap to say that because a couple was infertile the child will be too.

**Scott Gelfand, director of the Ethics Center at Oklahoma State University in Stillwater**

There is some research aiming to increase embryo survival and the likelihood that IVF will work. There are also people who are working on the other end — at the moment babies can only survive from around 22 weeks, but in future fetuses this could be extended to those that are 12 weeks. Someone could join these two advances together and we could have complete ectogenesis [in which the fetus develops outside the body in an artificial uterus]. I find it interesting and scary.

Those who work on artificial-womb technology aren't talking openly about it anymore. My guess is it's a potential lightning rod in our culture. There are some very interesting moral and ethical implications associated with artificial wombs. Certain conservatives might think it shouldn't be used but some might think that it could meet the test of *Roe vs Wade* — that it protects the privacy of the woman while preserving the rights of the fetus. If an artificial womb were developed,

the government could pass a law that requires people who have a termination of pregnancy to put the fetus into one of these wombs. That's the fear of many pro-choice theorists. There are

around 1 million abortions per year in the United States and there would have to be labs throughout the country, but if we put all these in artificial wombs and then put them up for adoption we would have one million more babies. It would be a nightmare. When I talk to some anti-abortionists about that, they really shudder.

Even in terms of insurance: if it became economically competitive with other forms of gestation, insurers might compel a person to use it to avoid premature birth or fetal alcohol syndrome. It's something that really needs to be talked about. Will it happen? Dolly was a complete surprise to everybody ...

**Miodrag Stojkovic, stem-cell biologist at the Prince Philip Centre of Investigation in Valencia, Spain**

Will we see a cloned baby? It could happen any day because of a lack of regulation [in some countries]. To my knowledge people are already trying to do reproductive cloning. Technically it is possible — a paper this year showed that you could derive human blastocysts after nuclear transfer [in which the nucleus of one cell is transferred to another cell] and our previous work has shown it is possible to grow human embryos to the blastocyst stage too. The only problem is the logistics of getting hold of enough viable human oocytes [eggs]. If we can make human oocytes from stem cells, it might be easier.

The field is developing so fast that some people can't follow what happens and are scared. There are plenty of movies in which scientists clone humans and use their organs for tissue donation, but this is not the reality. There is no medical need to clone a human. If we can make artificial gametes there would be no need to do it. The future is not about reproductive cloning, that's a very, very detrimental technique.

Humans are getting more and more lazy when it comes to reproduction. Male fertility is

declining and parents are deciding to have their first child at 40.

We need to learn what the embryo needs to result in a pregnancy — you can have excellent embryos in IVF and no pregnancy. But this can only be answered with human embryos. There is plenty we don't understand about embryo-mother communication. I don't think 30 years will be enough to answer those questions.

**Zev Rosenwaks, director of the Center for Reproductive Medicine and Infertility in New York**

I see the technology going towards possible eradication of infertility altogether. With nuclear-transfer technology or cell modification, I think we'll be able to generate sperm and eggs for anybody.

I think we've potentially reached the limit of biology in terms of the female's eggs, so artificial gametes might overcome that. In the best cases, success rates for IVF can exceed 70% with one cycle. The limitation is that we can't ensure the embryo is normal. I think there will be methods to evaluate the embryo for its viability and genetic competence in a non-invasive way, without having to remove cells. Cell-culture media is one way, as are micro-imaging techniques that look at chromosomal make-up. It could increase the success rate of IVF if we were better able to select the best embryo with markers such as metabolomics or with imaging techniques we can't even imagine today.

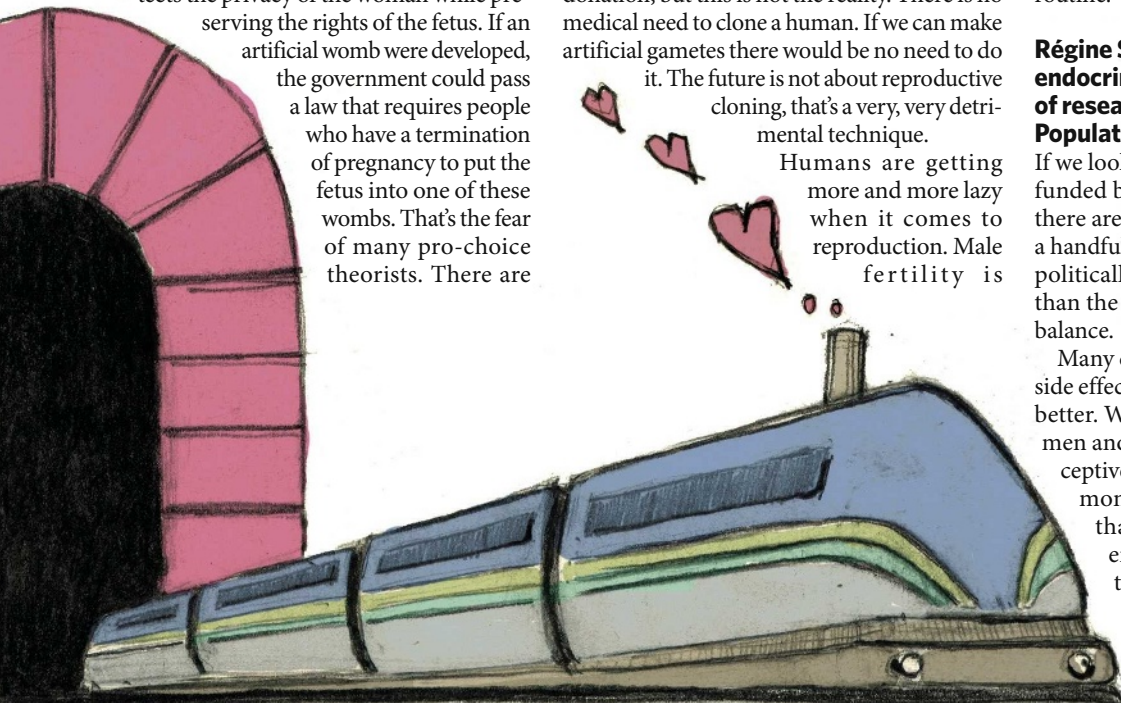
We can already take a cell from an embryo to create a cell line. In the future, this may become routine.

**Régine Sitruk-Ware, reproductive endocrinologist and executive director of research and development at the Population Council in New York**

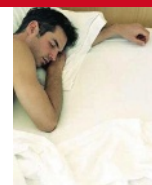
If we look at centres in reproductive sciences funded by the National Institutes of Health, there are more than twenty on IVF and only a handful on contraceptive research. It's more politically correct to help people get babies than the reverse, but it's important to have a balance.

Many current contraceptive methods have side effects or they're not effective. We can do better. We're hoping it might be possible for men and women to alternate taking contraceptives and that we can develop non-hormonal methods with fewer side effects that are very specific in targeting an enzyme or protein in the reproductive process, such as one that stops the ovum from maturing, or sperm from entering the egg. ■

See Editorial, page 253, and Essay, page 280.







**SLEEPLESS NIGHTS LEAD TO FALSE MEMORIES**  
But caffeine comes to the rescue ...  
[www.nature.com/news](http://www.nature.com/news)

PUNCHSTOCK

# Irrawaddy may be poisoned by arsenic

Drinking water in Myanmar's Irrawaddy delta may be contaminated with high levels of arsenic, a geological map suggests.

Arsenic contamination is a serious problem in several regions of south Asia, most notably Bangladesh, where half of the country's wells may be polluted. Long-term exposure to arsenic can cause skin lesions and cancer.

A new map of southeast Asia by geologist Michael Berg, of the Swiss Federal Institute of Aquatic Science and Technology in Dübendorf, and his colleagues shows the probability of arsenic contamination in the groundwater, based on geological and soil properties (L. Winkel *et al.* *Nature Geosci.* doi:10.1038/ngeo254; 2008). The region is particularly prone to arsenic contamination, Berg says, because of the abundance of large deltas filled with relatively young sediment deposited just 10,000 years ago. Young sediment is more reactive, and so thought more prone to releasing arsenic.

The map marks several recognized arsenic hotspots, including the Red River and Mekong deltas and the Chao Phraya basin. And it reveals a high probability of arsenic in excess of safe limits in southern Sumatra and the Irrawaddy delta — regions for which few data previously existed.

The models will spur debate because not everyone agrees that soil characteristics accurately reflect the sediment below. Years of weathering could have taken its toll on surface properties, argues John McArthur, a geochemist at University College London. "It's like asking you to tell me what's in the basement of your building by looking at the floor plan

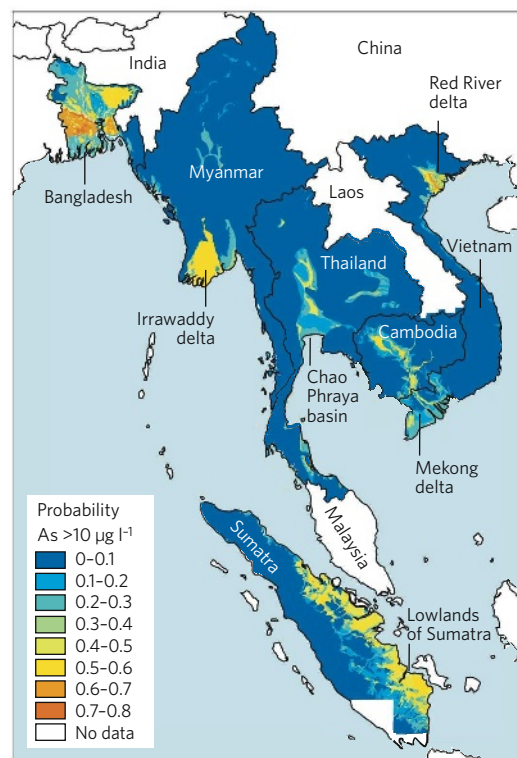


Skin lesions from arsenic poisoning.

of your lobby," he says. But Charles Harvey of the Massachusetts Institute of Technology in Cambridge notes that the sediment's youthful age means the surface may still hold useful information about the composition below. Berg and his collaborators found that their predictions matched known data 60–75% of the time.

Geologists agree that it would not be surprising to find arsenic in the Irrawaddy delta. A 2002 survey of 25 villages in the region found high levels of arsenic in 67% of the wells examined. That survey covered only a fraction of the 3.5 million inhabitants of the region, Berg notes: "There is an urgent need to have a closer look there."

Because the arsenic originates in sediment below the surface, Cyclone Nargis, which hit



M. SAMPSON, RESOURCE DEV. INT.-CAMBODIA

the area in May, is unlikely to have greatly affected the region's groundwater. The inhabitants of the delta now mainly rely on rainwater as a source of drinking water, says Paul Jawor, a relief worker for Médecins sans Frontières who has recently returned from the region. The wells in the delta are currently mostly used for cleaning and washing, he says.

Heidi Ledford

## Starting small but adding up: a free maths archive

A small group of researchers is meeting in Birmingham, UK, later this month to plan a free digital library of mathematics.

All the mathematical literature ever published runs to more than 50 million pages, with around 75,000 articles added each year. Over the past decade there have been several attempts to make this prodigious body of work accessible in a single digital archive, but so far none has succeeded.

A group of mathematicians

intends to change this. They have started small, with a handful of digitization projects in Poland, Russia, Serbia and the Czech Republic. In a few years they hope to unite these repositories with their western European counterparts in an archive to be hosted by the European Union, according to the organizer, Petr Sojka, an informatics scientist at Masaryk University in Brno in the Czech Republic. Eventually this pan-European archive could be expanded globally, he says.

To make such an archive easier to search, researchers have found ways to guess the subject of a paper on the basis of the frequency of symbols in it. But there will be many more-practical challenges, such as finding the funds to scan millions of old papers and striking deals with publishers who hold rights to them.

It may already be too late to build a single free mathematical archive, according to John Ewing, head of the American Mathematical Society, which maintains a list of more than

1,500 journals whose archives have already been digitized. "A few years ago, this model had the potential to change the mathematics journal literature in profound ways," he says. But most publishers have rushed to scan their own archives in order to lock them up and sell them to libraries.

"While the effort to digitize the smaller collections is admirable, and it's certainly worthwhile, it's unlikely to effect a larger change," says Ewing.

Jascha Hoffman



## WORD WATCH

## Black hole

A county commissioner in Dallas who compared the local traffic-ticket collection office to a "black hole", was told by a judge last week to apologize for his "racist remark". He had to explain that it was a scientific term.

## SCORECARD



## Sausage dogs

A survey of 6,000 dog owners and 33 breeds found the dachshund to be most aggressive. Their nastiness may have gone unnoticed because of their diminutive size, the researchers say.



## Dog sausage

China has ordered dog meat to be taken off the menu at its official Olympic restaurants, to avoid offending foreign visitors.

## SHOWBIZ NEWS

## Lab rats

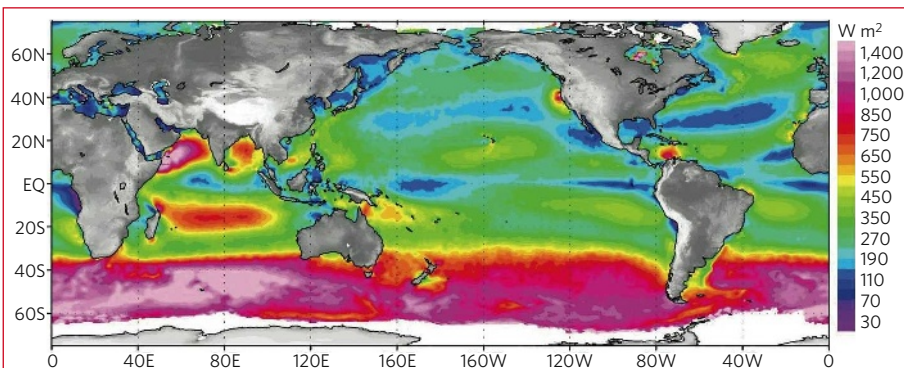
Following in the steps of US TV's *The Big Bang Theory*, comes a new British sitcom *Lab Rats*, which revolves around a fictional lab at St Dunstan's University and the comic incompetence of its staff (led by a drunken professor living off the reputation of his ageing Nobel prize). Though dubbed a "new formula" for comedy, sadly the reviews have largely called this TV experiment a failure.

## ZOO NEWS

## Bear suit

The legal owner of Knut, the now-famous polar bear that was rejected by his mother and raised by keepers at Berlin Zoo, is going to court. The Tierpark Neumünster says it wants a share of Berlin Zoo's increased profits, and has vaguely threatened to remove Knut from Berlin if it doesn't get its way.

Sources: Dallas Morning News, Appl. Anim. Behav. Sci., Xinhua, BBC, Bloomberg



## SNAPSHOT

## Global wind power

This map of the average power of the world's winds is derived from 8 years of data from a space-based radar called a scatterometer. The radar measures the scatter in microwaves bouncing off the sea into orbit; the wave motion measured indicates wind speed. The image shows wind-power density in December, January and February; during June, July and August the density becomes greatest

in the high latitudes of the Southern Hemisphere.

Many areas with high wind-power values experience mid-latitude winter storms, trade winds or monsoons. In other cases, the shape of the land funnels the wind into reliable jets. Timothy Liu, a researcher at the Jet Propulsion Laboratory in Pasadena, California, who compiled the map with his colleagues (W. T. Liu *et al.* *Geophys. Res. Lett.* 35, L13808;

2008), says that the higher producing areas could offer 500–800 watts of energy per square metre. But even with perfect turbines, only 60% of that could be recovered. And given how far apart turbines need to be, realistic numbers would be much lower still.

Europe's wind power capacity is estimated to swell from 56.5 gigawatts now to up to 40 gigawatts offshore by 2020. The United States is far behind, but projects in Delaware and Massachusetts are finally moving ahead. ■

Emma Marris

# Mouse miRNA library to open

Biologists working with 'knockout' mice that have been genetically engineered to inactivate certain genes are creating a library of knockouts of all microRNAs (miRNA) known in mice.

The collection, an initiative of the Sanger Institute in Cambridge, UK, is the first to tackle the nearly 500 miRNAs that scientists have identified in the mouse genome. The resource will eventually be available to all researchers studying the diverse systems affected by these short strands of nucleotides that regulate gene expression. Scientists speculate that mice may have more than a thousand miRNAs.

The Sanger project, named  $\mu$ KOMP because it's a micro counterpart to KOMP, the Knockout Mouse Project for protein-coding genes run by the US National Institutes of Health, should help those who are making knockout mice. It will offer the genetic tools and, ultimately, embryonic stem-cell lines for knockout mice.

"This will actually cut out a part of the process," says Allan Bradley, director of the Sanger Institute. Making a knockout mouse can soak up a year or so of a researcher's time, and cost thousands of pounds with no guarantee of success.

"I wish they had it already," says Ying-Hui Fu, a neuroscientist at the University of California, San Francisco, who wanted to knock out an miRNA gene for her work on circadian rhythms and multiple sclerosis. "More and more, people are realizing almost every biological pathway has miRNA involvement."

miRNAs regulate protein production in several biological systems and have been implicated in cancer and heart disease. In 2007, four labs simultaneously published reports of the first miRNA knockout mice<sup>1–4</sup>. Depending on the miRNA deleted, these mice develop immune deficiencies or holes in their hearts.

The Sanger team will make all the knockouts using one method, and in the same strain of mice, so scientists will be able to directly compare different mutations. "That is so important," Fu says. "Different strain backgrounds sometimes give you totally different results." ■

## Amber Dance

1. van Rooij, E. *et al.* *Science* **316**, 575–579 (2007).
2. Thai, T.-H. *et al.* *Science* **316**, 604–608 (2007).
3. Rodriguez, A. *et al.* *Science* **316**, 608–611 (2007).
4. Zhao, Y. *et al.* *Cell* **129**, 303–317 (2007).

## Court victory for subject of experimental therapy

Dieter Bollmann of Berlin, who underwent unsuccessful stem-cell therapy for urinary incontinence, has won a court case against authorities of the University Hospital Innsbruck in Austria. He had claimed that the urologists who treated him had not told him the procedure was experimental.

On 11 July, a civil court awarded Bollmann nearly €3,000 (US\$4,800) in damages and expenses. His lawyer is considering claims from several other former patients who were similarly operated on within the past few years.

The therapy involves generating muscle stem cells from a biopsy taken from a patient's arm, then injecting them into the same patient's urinary sphincter in an attempt to boost its contractile power. The urologists are also involved in an independent dispute with their local ethics committee, which claims they carried out clinical trials — one of which was published in *The Lancet* — without ethics approval (see *Nature* 453, 6–7 and 1177; 2008).

## Drug firms join forces to boost discovery pipeline

Pharmaceutical giants Merck, Pfizer and Eli Lilly have joined forces in a start-up company that they hope will stimulate drug discovery.

Boston-based Enlight Biosciences will develop lab technologies to the 'pre-competitive' stage, so that each company will be able to use them to take its own research forward. The firm's leaders include a coalition of drug-development and academic

researchers, such as Robert Horvitz, a Nobel laureate biologist at the Massachusetts Institute of Technology in Cambridge.

Enlight will focus on areas such as imaging technologies, toxicology studies and speeding up the synthesis of interesting compounds. Relatively little money has been committed so far — just \$39 million has been announced to target these areas and many more.

## Chemical giant seeks route into speciality market

US chemical company Dow is buying out rivals Rohm and Haas for more than \$15 billion.

The move is part of an attempt by Dow, which is headquartered in Midland, Michigan, to move into speciality chemicals and away from the bulk products that have traditionally made up its business. The acquisition of Rohm and Haas, based in Philadelphia, Pennsylvania, would make Dow the largest speciality chemicals firm in the United States, according to company officials.

Dow will pay \$78 per Rohm and Haas share, a substantial premium on stock that was trading at around \$45 before the announcement.

## 'Fundamentally flawed' US pollution law thrown out

A US appeals court struck down landmark air-pollution regulations last week, shocking both environmental and industry groups with a decision that could severely hamper efforts to curb smog and acid rain.



Efforts to curb pollution in the United States have suffered a legal set-back.

The 2005 Clean Air Interstate Rule created a market-based programme allowing utilities to buy and sell permits in order to meet new limits set for emissions of sulphur dioxide and nitrogen oxides. Many within the utility industry supported the programme, but North Carolina and a number of power companies challenged various aspects of the law, including how the permits were distributed. The court agreed, calling the programme "fundamentally flawed" and ordering the Environmental Protection Agency (EPA) to start anew.

The decision came on the same day that the EPA said it would delay issuing the regulations on how to deal with carbon dioxide emissions that the Supreme Court had demanded.

For a longer version of this story, see <http://tinyurl.com/6pafqx>

## FDA aims for neutral tone in drug-rejection letters

In a move intended to stop investors from over-interpreting its decisions, the US Food and Drug Administration (FDA) will next month adopt a new system for delivering less-than-stellar news to companies trying to bring new drugs to market.

From 11 August, the agency will stop sending out its current genre of missives, which come in two categories: 'approvable' letters, if the drug needs more information or specific changes to win approval, and 'not approvable' letters, for applications with more major problems. In their place, a company whose drug is being rejected will receive a 'complete response' letter outlining an application's deficiencies and what can be done to address them.

Because the letters are private, investors will be able to infer little from a complete response except that a drug hasn't yet met the FDA's marketing standards. The change, says Janet Woodcock, the agency's top drug-approval official, will help the FDA to adopt "a more consistent and neutral way" of delivering the news that a drug application is not up to standard.

## Lionfish not a roaring success for coral reefs

The red lionfish (*Pterois volitans*), a stunning Indo-Pacific species, is central to an invasive-species worst-case scenario that is playing out in the Atlantic Ocean. A study suggests that lionfish, with their voracious appetites, can reduce the survival of young fish on experimental reefs by 80%.

Eradicating the lionfish is now all but impossible in the Bahamas, warn the study's researchers. But conservationists in nearby regions in which lionfish have been spotted in smaller numbers, such as the Dominican Republic, could take steps to guard against the interlopers while there is still time.

The study was presented last week at an international corals conference in Fort Lauderdale, Florida.



M. ALBINS

D. GRAY/REUTERS





# THE LONG SUMMER BEGINS

A research vessel embedded in the thinning Arctic sea ice has a front-row seat for the cryospheric show of the century. **Quirin Schiermeier** reports from Darnley Bay, Canada.

It is a mild, sunny day in June. Off the coast of far northern Canada, the icebreaker *Amundsen* is having little trouble ploughing through the land-fast ice. The sky is azure — the ice beneath it is rapidly melting and breaking up. “I’ve been doing research in the Arctic for 25 years now, but I’ve never seen such a long period of clear skies and mild weather,” says Dave Barber, lead scientist for the sea-ice study under way, looking out from the *Amundsen*’s bridge.

Only a few days earlier, on this same spot in Darnley Bay, scientists had deployed monitoring equipment on what seemed like stable ice. When they returned, the ice had broken up into a jigsaw puzzle of myriad floes. On one, a meteorological tower was drifting precariously close to the edge. Barber sent out a flat-bottomed ‘skippy’ ice boat (pictured above) to retrieve the precious station, then dispatched a helicopter to search for other instruments that had been scattered adrift on other floes. Remarkably, all were retrieved within an hour.

The *Amundsen* sits in just one tiny area of the vast patchwork of floating ice that covers much of the Arctic Ocean. But it happens to provide a first-hand look at a scientific question with

a peculiarly iconic importance: will this summer break last year’s record low in the extent of Arctic ice cover? Last September, at the end of the melt season, the sea-ice area measured just 4.3 million square kilometres, 39% less than the long-term average for 1979–2000. Coming on top of a long-term trend towards shrinking sea ice<sup>1</sup>, this observation prompted ever-more-gloomy predictions that the Arctic summers will become entirely ice-free — perhaps as early as five years from now.

So far, the 2008 sea-ice cover is not dramatically below average (see map). In fact, at the end of June, ice covered 7.8 million square kilometres of the Arctic Ocean, compared with 7.5 million square kilometres at the same time last year. The difference is that much of the ice this year is first-year ice — frozen only last winter — and thus much thinner and more prone to melting than ice built up year on year.

The factors contributing to the record-low ice cover in 2007 included a high-pressure weather system that settled in over the

Beaufort Sea and a low-pressure system over Siberia; strong winds persisted between the two, which helped to shift around and melt the ice throughout the summer<sup>2</sup>. Whether similar weather patterns will come into play during July and August this year remains to be seen, says Mark Serreze, a sea-ice expert at the University of Colorado at Boulder.

*In situ* measurements such as the *Amundsen*’s are crucial as the Arctic is changing so rapidly that some of the old rules about how sea ice behaves don’t hold true any more. “Winter processes, such as winter ice growth rates, are key for understanding what’s going on in the Arctic —

but direct oceanographic measurements are sparse and very hard to come by in this harsh environment,” Serreze says. “Having folks up there during the cold and dark winter is the only way to do it.”

Captain Lise Marchand and her crew know about the cold and dark. They brought the ship, a converted Canadian Coast Guard vessel, into the southern Beaufort Sea last October. Throughout the autumn, winter

**“There is no doubt that the Arctic is moving in a trajectory for becoming a new environment.”**

— Dave Barber



and spring, at times locked in thick pack ice, at times in open water, the scientists on board probed almost everything that could be sampled in this pristine environment: from sea ice to marine life, from sea-floor sediments to atmospheric gases.

The 400-person, US\$40-million expedition is the biggest single project within the International Polar Year of 2007–09. Its goal is to sit as long as possible within the circumpolar ‘flaw lead’, a semi-permanent area of open water within the ice pack. The *Amundsen* is the first icebreaker to overwinter in a flaw lead, a fact that Barber is quite proud of. “Nobody, not even the Russians, has done anything like this before,” he says.

### Holes in the ice

Flaw leads are a recurrent peculiarity all around the Arctic basin, ripped open as winds push the thin skin of sea ice to rotate around its centre of mass near the North Pole (see video, <http://tinyurl.com/6zkgxm>). “Just think of flaw leads as areas where normally there should be ice, but where there is none,” says Barber, of the University of Manitoba in Winnipeg. The Amundsen Gulf, where the *Amundsen* works, is one such place. It usually hosts a flaw lead of some sort, although it is not always ice-free in the winter.

From its front-row position in the flaw-lead system, the ice-breaker is well placed to observe the many interactions that lead to changing sea-ice conditions. If the ice shrinks earlier in the season and moves faster, flaw leads open more often and possibly in different places from before. As a result, a larger area of water will be exposed to sunlight earlier in the year than usual.

The *Amundsen* expedition, known as the Circumpolar Flaw Lead System Study, is here in part to pin down details of the ice–ocean albedo: the fact that the dark surface of the open ocean absorbs more sunlight than does white snow and ice. Although it might not be the full story<sup>3</sup>, the ice–ocean albedo is widely invoked to explain why the Arctic has warmed about 3.5 °C since pre-industrial

times — three to four times more rapidly than other regions. Studies being done on the *Amundsen* are meant to illuminate details such as how the optical properties of the sea ice — including melt ponds atop the ice itself — alter the albedo.

Other, more dynamic effects are even less well understood. As ice melts earlier in the season, the atmospheric dynamics change because of heat released by the open water. Warmer waters begin to enter the Arctic basin and storms can form that blow massive amounts of ice out of the basin through the straits west of Greenland. Less ice in the basin means that the remaining ice pack starts to spin faster and faster, and to reverse its rotation more frequently.

What this all means for long-term sea-ice cover remains to be seen. Most computer models have not managed to predict anything like the massive ice losses observed in the past few years. So some think that their predictions of a slow decline in sea ice over the coming century is a massive under-estimate. “Over the next five to ten years we are almost certainly going to lose all the sea ice” in summer,

**“In my opinion there is a fifty-fifty chance of a record melt this year.” — Bill Chapman**

climatologist James Hansen, of the Goddard Institute for Space Studies in New York, told politicians in Washington DC last month.

Bets have already been laid on whether this summer’s ice loss will be more than last year’s. William Connolley, a software engineer who used to model sea-ice changes at the British Antarctic Survey, has taken in roughly €300 (US\$470) so far in the informal online pool he runs. (He bet ‘no’.) Connolley points out that, even if a new record is

set, that has little meaning in the long term. “We all recognize the climatological trend is downwards,” he says, “but what will happen this year depends on weather and natural variation.” As for recent media speculation that the ice at the North Pole will be melted this summer, he notes: “There may or may not be open water in the area by September, but nobody is predicting there will be zero sea ice in the Arctic, or that sea-ice extent might halve compared with last year.” The possibility of an ice-free North Pole, he says, is a statistical fluke about a particular geographical point that says little about sea-ice conditions in the basin as a whole.

“In my opinion there is a fifty-fifty chance of a record melt this year,” says Bill Chapman, an atmospheric scientist at the University of Illinois at Urbana-Champaign. Chapman runs the *Cryosphere Today* website (<http://arctic.atmos.uiuc.edu/cryosphere>), which compiles data from satellites to measure the extent of ice and open water on a daily basis.

### Daily perils

From the perspective of researchers aboard the *Amundsen*, the ever-shifting sea ice is more than just a long-term trend to ponder; it is also a daily danger to the ship’s very existence. In March, the team had chosen an ice bridge between Cape Parry, on the Canadian mainland, and Banks Island as a place to set up an ice camp. But this plan was never implemented. The ice bridge stayed in place for only five days, and when it broke strong winds began to blow the sea ice behind it into the Beaufort Sea, nearly crashing the



DOUG BARBER

*Amundsen* into Banks Island. Barber says that the break-up of this ice bridge is very unusual, as was the absence of ice in the Amundsen Gulf from October to December last year. This could have been, he says, a knock-on effect of the unusual ice loss last summer. "There's an accumulation of effects," says Barber. "What's happening in one year will affect the ice in the next."

If there is a lot of open water in late summer, as there was in 2007, the heat that has accumulated in the ocean's upper layers takes longer to dissipate, delaying the onset of new ice even if air temperatures are very low. (Last year, ice didn't form until quite late in the year.) In addition, the presence of open water means that sunlight penetrates the water for longer, deepening and warming the uppermost layers. Together, Barber says, these feedbacks almost always mean that rapid ice loss — such as that seen in 2007 — will lead to a reduction in sea ice the following year, unless extremely cold temperatures intervene.

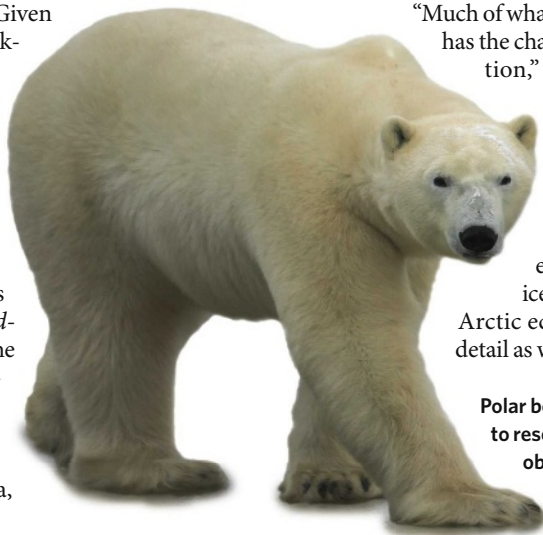
### Seasonal colours

In late spring the Arctic is a mixture of white ice, green-blue melt ponds, and various transition zones in between. Understanding the seasonal evolution of the different surface types, and their respective contributions to the albedo, is important for calculating the energy balance of a given area, and of the Arctic Ocean as a whole. Knowing how the thin land-fast ice changes, breaks up and melts during spring will also make the interpretation of radar satellite observations easier. Radar sensors cannot easily distinguish between the different surface features of sea ice without this sort of 'ground-truthing'.

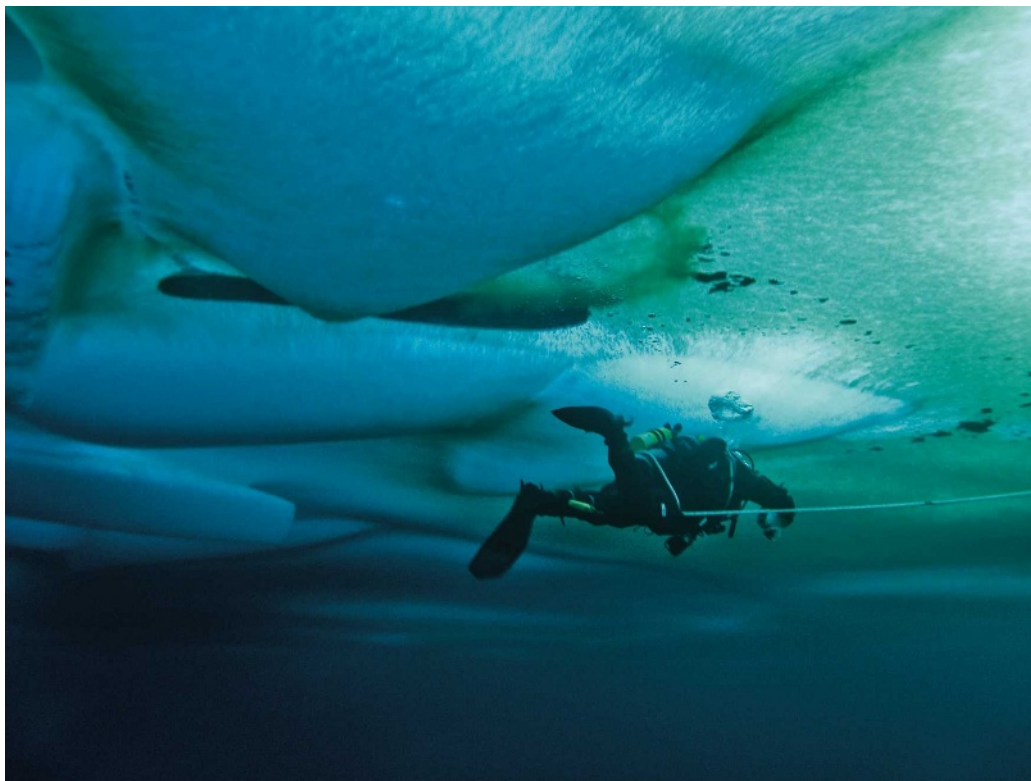
The Circumpolar Flaw Lead System Study is working to quantify not just what factors change the sea-ice melting, but also what happens to the entire environment when

flaw leads open up. Given the accelerating break-up of sea ice, many scientists think that the amount of light, nutrients and food available to Arctic ecosystems — from algae to polar bears — will change dramatically. Researchers on board the *Amundsen* are performing the most extensive exercise ever in taking stock of the Arctic environment and ecosystems. The data,

T. J. RICH/NATUREPL.COM



Polar bears pose a danger to researchers trying to obtain samples from the ice sheet.



Divers on the *Amundsen* explore the marine environment, where thinning ice has created pockets for animals to live and hide in.

they hope, can then be used as a baseline against which future changes can be compared.

For the scientists on board, most of whom are master's and PhD students, this overarching goal translates first and foremost into tireless sampling and measuring from the ocean, ice and atmosphere. Some samples are processed at sea, in the *Amundsen*'s 12 interior laboratories and cold rooms. Others will be sent to labs around the world for more detailed analysis.

"Much of what we're doing here has the character of exploration," says Jens Ehn of the University of Manitoba, who measures the optical properties of sea ice. "Nobody has ever looked at the ice and its role in the Arctic ecosystem in such detail as we are."

**"You never really know what you will find."**

— Jens Ehn

Changes in the seasonal evolution of sea ice are thought to have lasting effects on primary productivity — the production of organic compounds through processes such as photosynthesis — and the composition of species that live beneath the ice. Studying these changes requires scientists from different disciplines to work together. Life aboard the *Amundsen*, where scientists who would usually speak with each other only rarely now meet every

day — in the mess, in the lounges, at the daily science meeting, or during the bar nights every Tuesday, Thursday and Saturday — facilitates such multidisciplinary work.

### Light work

Ehn, for example, uses a spectroradiometer to measure how much light goes through the different surface types and reaches the algae communities in the water column. At the same time a couple of scuba divers, including Haakon Hop, a marine biologist with the Norwegian Polar Institute in Tromsø, collect algae samples while exploring the ice from below. The researchers may then conclude a busy day discussing ideas and observations with Christopher Mundy, a biologist at the University of Quebec in Rimouski who studies species



succession and habitat evolution in Arctic environments.

As the ice melts, a plume of fresh water forms right beneath it. How algal communities adapt to this fresh water, and how primary productivity switches from the ice–water interface to greater depths, is what Mundy's team is trying to find out. "Seasonal changes in species composition beneath the ice haven't been studied to a large extent, mainly because diving here is dangerous and logistically challenging," he says. There are no results yet, but the researchers are already seeing that the many small holes and domes that form in the ice when it starts to melt become temporary habitats for organisms to escape predation. The scientists are now looking at how the ecosystem responds as these habitats disappear. "You just have to be at the right places at the right time to study these things," says Ehn. "Even so, you never really know what you will find."

### Out and about

Sampling on the ice is a tough job in cold weather, particularly during the long polar night, when even brief exposure can cause frostbite. Operating instruments or typing on laptop computers can be a real challenge then. But when the weather is as mild and friendly as it was in early June, a few hours' work on the ice — in heavy rubber boots, special 'Mustang' suits and with one team member carrying a gun (as protection against polar bears) — are a welcome change. For the scientists on board the *Amundsen* this is the only opportunity to leave the confined space of the ship that can quickly become an all-too-familiar living space.

Even so, most researchers find it thrilling to spend up to 12 weeks on board an icebreaker cramped with first-rate science equipment and colleagues from many disciplines. "Being here is really an extra motivation for me," says Eva Alou, a PhD student at the University of Quebec, who spends up to 15 hours per day in a small lab measuring photoactivity in zooplankton cells. She is trying to find out how exposure to ultraviolet light affects the cells' repair systems, which are known to work slowly in the cold Arctic environment. Many

marine scientists consider changes happening at the low end of the food web — in microorganisms and algae — to be more critical than those that affect iconic species such as whales and polar bears.

But birds and mammals (including humans, as indigenous people are another line of study) are also a focus. Natalie Asselin, who studies environmental geography at the University of Manitoba, is doing an aerial survey of beluga whales, which use the flaw leads as feeding and breathing grounds during spring. Changes in the distribution of the sea ice, which shelters the whales from waves and from killer whales, are thought to affect the population in some

thanks to the winter data," says Yves Gratton, an oceanographer at the National Institute for Scientific Research in Quebec.

Gratton leads the study's physical-oceanography team, which is investigating what will happen to the larger-scale ocean circulation if the ice disappears. Warm water is known to be entering the Arctic Ocean through the Fram Strait; that warmer and denser Atlantic water is thought to accumulate at the bottom of the Arctic basin. There are worries that if this warm water reaches the surface, via small-scale ocean mixing processes, it might further accelerate Arctic warming and sea-ice loss.

### Mix and match

Gratton and his team use CTD data to determine the vertical structure of ocean water, and to extrapolate and model turbulence and mixing processes. With instruments that measure temperature at a millimetre scale, they hope to be able to quantify the amount of ocean mixing in the upper 100 metres. Data analysed so far have revealed two eddies, both around 20 kilometres in diameter, that travelled across the Beaufort Sea in February and March. The scientists will now investigate whether these eddies might be a regular feature — and, if so, what that might mean for the environment around them.

And beyond, "there is no doubt that the Arctic is moving in a trajectory for becoming a new environment", says Barber — changes that will "sooner or later communicate to the whole planet". In the end, the *Amundsen's* front-row seat in the new Arctic theatre may be a glimpse of the show that will unfold for us all. ■

**Quirin Schiermeier writes for *Nature* from Munich.**

1. Parkinson, C. L. & Cavalieri, D. J. *J. Geophys. Res.* **113**, C07003 (2008).
2. Zhang, J., Lindsay, R., Steele, M. & Schweiger, A. *Geophys. Res. Lett.* **35**, L11505 (2008).
3. Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E. & Svensson, G. *Nature* **451**, 53–56 (2008).

**To read Quirin's trip diary from aboard the *Amundsen*, visit [http://blogs.nature.com/news/blog/events/canadian\\_icebreaker](http://blogs.nature.com/news/blog/events/canadian_icebreaker).**



Members of the sea-ice team drill and analyse cores to look into the region's history.



way or another. How much the whales will be affected by the ice loss is hard to foresee at this point, she says.

The amount of data that will have been collected by the end of the project is impressive. One sampler device, by itself, will yield some 1,500 profiles of water conductivity, temperature and depth. Called CTD profiles, these data can help oceanographers to compute the density and salinity of ocean water. A plethora of data is also being collected from moored observatories, free-drifting buoys, current meters, sediment traps, echo sounders, an on-board gas-flux measurement tower, a tethered underwater vehicle and scuba divers — and this list is far from complete. "We'll end up with an absolutely unique data set, in particular



# FROM THE DESERT TO THE EDGE OF SPACE

Not all NASA launches need rockets and countdowns.  
**Eric Hand** sees the alternative in Fort Sumner, New Mexico.

**O**n the day of the launch Mark Cobble arrives well before dawn. Lightning studs the sky on the horizon; halogen spotlights bathe the three-story NASA hangar in light. The desert around Fort Sumner municipal airport in New Mexico is still, its silence relieved only by the hum of a generator, the clapping of an exhaust flap on a helium truck and the slight rustle of the wind in the scraggly trees and scrub. That windy rustle bothers Cobble, the NASA crew chief in charge of the launch that is about to get under way.

Since 02:30 the airport manager has been sending up weather balloons the size of beach balls every half hour or so. Ross Hays, a former CNN weatherman now with NASA, has been on site since 01:00 to gather weather data, armed with two flasks of coffee and the peanut butter and jelly sandwiches he reserves as a treat for 'show days'.

"No chance of precip?" Cobble asks.

"Nope," says Hays, an optimist. "It's all over in the panhandle. It's all clear over New Mexico."

Cobble looks at the weather data gleaned from the Air Force, the National Weather Service and a network of instruments at public high schools. He strokes his chin. The wind is definitely worrisome. But the weather is only going to worsen as

the week goes on. "We'll see how it goes and hope for the best," he finally says. "I'll get 'em started."

If your image of a NASA launch is a mission control room full of computers, pocket protectors and neck-ties, think again. This is country-and-western NASA, all boots and blue jeans; the tools of the trade are Buck knives, duct tape, parachute cord and patience, plus a whole lot of helium. And the 'spacecraft' hang beneath balloons.

Balloons have long offered a cheap alternative to rocketry. Even if by their very nature they cannot rise above the atmosphere, they can still rise above 99% of it, leaving dust, weather and water vapour behind. In the 1980s, NASA was launching 80 balloons a year as part of its programmes in astronomy and atmospheric chemistry: balloon mounted telescopes weighing a tonne or more have made some remarkable discoveries — and remarkable recoveries. Unlike their brethren on satellites, sensors that fly beneath balloons can be used again and again, although getting them back from wind-blown landing sites can be a trial.

Since the 1980s, though, the number of NASA balloons launched has steadily declined. Last

year it was just 17. This is partly explained by the changing nature of the experiments and observations that astrophysicists want to make. Those that could be performed easily and quickly — the low-hanging fruit for telescopes 30 kilometres up in the sky — have been done. Now, most experiments require longer flights and more complicated instruments. And this is where a

new sort of balloon comes in — the peculiarly pumpkin-shaped balloon folded up and wrapped in protective red plastic in the Fort Sumner hangar, which Cobble and his crew from NASA's Columbia Scientific Balloon Facility (CSBF) are about to

launch. Its fabric, as thin as the plastic of a refuse bag, will, when inflated and aloft, measure 54 metres across and have the volume of more than 600 double-decker London buses. And if all goes well, it will provide a solution to the abiding problem of scientific ballooning — that what goes up comes down all too soon. According to its designers, this style of a balloon can stay up night and day for months at a time.

**"They have to prove the technology before people start to take them seriously."**

— Simon Swordy

## Getting high

All scientific balloons are filled with helium. In the more-normal balloons, sometimes called zero-pressure balloons, this buoyant helium is at the same pressure as the atmosphere outside, and the balloon's volume thus changes with the temperature. This means that if it is to retain its altitude, the balloon may need to vent gas by day and drop ballast by night. This limits the mission lifetime, normally to days. In steady conditions such as the constant daylight of a polar summer balloons can stay up longer; in 2005 a NASA balloon circled Antarctica three times in a record 42-day flight. But although that is long enough for astrophysicists looking at things such as cosmic rays to get good data, longer would be even better. Immunity to the day-night cycle would also allow missions to fly from more accessible places than Antarctica, and would make things easier for observers who like to do it in the dark.

The balloon being launched at Fort Sumner is designed to provide these benefits by

## Where balloons have made advances

### Cosmic microwave background

In 1998, a project called Balloon Observations of Millimetric Extragalactic Radiation and Geophysics (BOOMERanG) made a fine-scale map of the cosmic microwave background over a part of the sky and provided evidence for its 'flatness'. A follow-up flight in 2003 was able to detect some polarization in the background — consistent with a period of 'inflation' in the very early cosmos.

### Antiparticles

The Balloon-borne Experiment with a Superconducting Spectrometer (BESS) made nine 1-day flights in 1993–2002 and detected more than 2,400 antiprotons, thus contributing 80% of the records of incoming antimatter. More recent longer flights revealed constraints on the presence of anti-helium. The instruments rival the sensitivity of the \$1.5-billion Alpha Magnetic Spectrometer — still awaiting a ride to the International Space Station.

### Supernova 1987a

When the most spectacular supernova in centuries exploded while America's space shuttles were grounded, a detector that might have flown on the shuttle was launched on a balloon from Antarctica. Although that pioneering effort failed, several balloon missions in Australia over the following two years detected rays from the explosion, confirming theories about the radioactive decay of the star's remains.

D. WAKEFIELD

maintaining a constant volume as opposed to constant pressure. The gas inside will at times be at a higher pressure than the atmosphere outside, which is why such designs are called 'super-pressure' balloons. The pressure changes are absorbed by high-strength tendons in the balloon's ribs made from Zylon, a polymer also used to protect racing-car cockpits. NASA scientists think super-pressurized balloons like this could stay aloft for 100 days or more (see graphic, overleaf). Because they are less vulnerable to diurnal cycles they also open up the possibility of multiple trips around Earth at mid-latitudes. That opens up more of the sky to observation, and lets strictly nocturnal astronomers into the game too.

### Outperforming satellites

More time, more sky; more night-time sky, more sky all year round. In terms of science per dollar, super-pressure balloons might well outperform satellites for a far wider range of missions than the current technology allows. "There are a lot of people waiting for this technology to become mature enough," says Simon Swordy, director of the Enrico Fermi Institute at the University of Chicago, Illinois, who wants to use balloons to study the history of cosmic rays in the Milky Way. "But they have to prove the technology before people in NASA start to take them seriously," he says.

Providing that proof is hard. The arguments in favour of super-pressure balloons have been around for a while; *Nature* reported on them enthusiastically five years ago (see *Nature* 421, 308–309; 2003). But flight success has proved elusive. The test at Fort Sumner is part of the latest attempt to do something about that. Failure could jeopardize an increase in NASA's planned budget for scientific balloons, which calls for spending on balloons to inflate by 50% over the next four years, in part so as to be ready to take advantage of super-pressure technology as it moves from understandably temperamental prototype to work horse. Too much temperament, though, and the programme will founder. The test flights have to go well. And the New Mexico weather has to cooperate with Cobble.

By 06:00 Cobble has his team hard at work. Big Bill, the 45-tonne truck that tethers the balloons until the last moment, lumbers out to a patch of cracked, weedy pavement. The sky has blossomed from black to blue to pink, and the rising sun reveals wind turbines ticking over on a distant ridge — not an encouraging sign for the balloon launchers. Danny Ball, Cobble's boss and the director of the CSBF, in Palestine, Texas, points to a small weather balloon on a long leash that the wind is pulling out at a 45° angle. "That's about 18 knots on it right now," Ball says. "That's way, way too much." While







### SATELLITES

**Cost:** More than \$10 million**Duration:** Years

Slow to build and expensive, satellites gather years worth of data with no atmospheric interference.

### SOUNDING ROCKETS

**Cost:** \$1.5 million**Duration:** 15 minutes

Quick up-and-down flights, to altitudes anywhere between 30 and 1,500 kilometres, provide training for young investigators, and explore regions too high for balloons and too low for satellites. But flights are very short.

### ZERO-PRESSURE BALLOONS

**Cost:** \$250,000**Duration:** Up to 30 days

Flying three times higher than aeroplanes, stratospheric balloons collect data above 99% of the obscuring in the atmosphere. Zero-pressure balloons contain an open valve. Daily temperature swings cause losses in altitude that can be compensated by ballast drops for several weeks. But long flights are limited to summers at the poles, where temperatures are nearly constant.

### SUPER-PRESSURE BALLOONS

**Cost:** \$1.5 million**Duration:** 100 days or more

The next generation design in scientific ballooning, super-pressure balloons are completely sealed. Tendons in the pumpkin-shaped balloon absorb pressure changes caused by temperature swings. The balloons dwarf a Boeing 747 aeroplane, but, as the indicators show, fly much higher. Flights could circle Earth at mid-latitudes for months; more of the Universe could thus be observed during both day and night.

Kilometres above sea level (log scale)

the launch team peels the balloon from its banana-skin wrapping and checks for tears and twists, Ball paces in a red-hooded sweat-shirt, shorts and sandals. He frequently cups his hands against the wind to light cigarettes that he then throws away half smoked.

All of NASA's balloon launches come under Ball's jurisdiction and he embodies the programme's blue-collar culture: gruff and given to a blunt common sense. Before ballooning he worked on offshore drilling barges in the Pacific Ocean. "The satellite scientists work in clean rooms with 10-year development cycles and unlimited reviews," he says, as the crew lays out the wrapped balloon behind Big Bill. "We don't have clean rooms. We don't do any of that"

### More bang for the buck

The difference is reflected in the price. Most of the satellites that hold, stabilize and power scientific instruments are made by aerospace companies such as Boeing or Lockheed Martin for tens of millions of dollars, at the cheapest. The same companies, and some others, sell rocket launches at similar prices. NASA's balloons, on the other hand, are glued together for US\$250,000 a time on the work tables of Aerostar, of Sulphur Springs, Texas, a facility that has at times sold left-over polyethylene film for refuse bags. The tougher skin and tendons that go into super-pressure balloons make them cost a fair bit more — about \$1.5 million — but nothing near the cost of a satellite launch. The most expensive parts of such missions are the instruments themselves, which are often exactly the same as those that will fly on the satellite missions they foreshadow. But below a balloon they can be flown in simple gondolas hanging from steel cables; no need for the thermal insulation or the hi-tech components in satellite buses. Ball even recalls flying an instrument still supported by the workbench on which it was built.

Price is not the only difference. Compared with the fairly fast turnaround culture of the balloons, satellite development is sluggish and risk averse. Instruments can fly on balloons years before their equivalents or descendants fly in space (see 'Where balloons have made advances'). In 1998, the Balloon Observations of Millimetric Extragalactic Radiation and Geophysics (BOOMERanG) experiment studied the cosmic microwave background in part of the sky, finding important evidence for the 'flatness' of the



Universe and scooping some of the results from the later and much more expensive Wilkinson Microwave Anisotropy Probe. Later this year, when the European Space Agency's €600-million (\$940-million) Planck spacecraft starts to map the whole-sky background it will be using instruments of the same design.

According to Andrew Lange, an astrophysicist at the California Institute of Technology in Pasadena who was on the BOOMERanG team, a balloon could fly something much more impressive today. He has plans for an instrument called SPI-

**"Some people have the perception: 'It's just a balloon. How hard can it be?' But it's a complicated structure."**

— Henry Cathey

DER that would measure the polarization of the microwave background, and thus reveal the work of gravitational waves in the early universe. If he were to get SPIDER onto one of the first operational launches of a super-pressurized balloon, as he hopes to in a few years'

time, he thinks such measurements might provide conclusive proof of cosmological inflation — and thus scoop a proposed but not-yet-funded space mission, the Inflation Probe, which is unlikely to see space much before the end of the next decade.

Academics welcome the rapid concept-to-completion arc of balloon experiments not just as a way of flying equipment sooner, but also as a ready-made system for minting PhDs. "The timescale of building a balloon instrument, flying it, and getting data, is a timescale that a grad student can see from beginning to end," says Martin Israel, an





astrophysicist at Washington University in St Louis who is chairing a scientific ballooning planning report for NASA due later this year. The importance of that fertile training ground is not to be underestimated, says Swordy, who laments the “armies” of NASA scientists who sit behind computers but know little about building instruments. “We have to preserve a cadre of people who know how to put together instruments for the future.”

### Lift-off

At 07:00, during a long lull in the breeze, everyone around the test balloon puts in earplugs. Cobble opens valves on the helium truck and with a high-pitched hiss the gas snakes its way up tubes and into a balloon. Not yet into the super-pressurized balloon, but into a tow balloon attached to the top which lifts the main balloon up. Then 900 cubic metres of helium is introduced into the main balloon, enough to give it the anti-gravitational equivalent of a small car's weight in lift. Twenty minutes later, Cobble, riding at the front of Big Bill, pulls a lever and releases the balloon from a spring-loaded pin. Two dozen weathered faces, with shaded eyes, follow it up until it disappears in the glare of the still-low sun (see ‘Up, up and away’).

Its charge discharged, the CSBF team relaxes. But the engineers who designed and built the balloon are still on tenterhooks. At ground level the balloon is far from fully expanded; it is supposed to take on its designed pumpkinoid shape only in the much lower pressure of the stratosphere. But its predecessors have failed to do so. Since 2001, something odd has been happening on the test flights; the balloons, as

they have risen, have never fully expanded. Material has remained folded in on itself in wrinkled clefts that look like Pac-Man or the S-shaped stitches on a baseball. That leads to instability of a sort that rules the balloons out for any practical work.

Structural engineers were called in to model the problem. They found that there was too much material in between the longitudinal tendons at the balloons' poles. As a result of this, the balloons naturally tended to fold in on themselves to reduce stress, says Henry Cathey, a NASA engineer at the Physical Science Laboratory of New Mexico State University in Las Cruces. “Some people have the perception: ‘It's just a balloon. How hard can it be?’ But it's a complicated structure.”

In the hangar, the engineers watch live video from two cameras pointed up at the underbelly of the balloon from the gondola below. Its posterior blossoms like a flower. As the balloon rises at 18 kilometres per hour, wrinkle after wrinkle disappears. Except for one, at two o'clock on the balloon's downturned face. The engineers urge the recalcitrant cleft on:

“Come on. Come on.”

“Give the dog a bone.”

Finally, just as the balloon reaches its target height of about 30 kilometres, the last cleft relaxes; relief spreads through the room as everyone claps and shakes hands. The balloon floats west for an hour before the team triggers a tear in one of its panels. The engineering payload and its parachute drop toward the Capitan Mountain Wilderness of the Lincoln National Forest, more than 100 kilometres away, and land safely, dangled over a barbed

wire fence and a rancher's trees — the balloon drifts on 20 kilometres downwind. The crucial goal of a fully puffed-up deployment has been achieved. In August, the balloon programme will ramp up for a bigger test: a super-pressure balloon the size of the Titanic (which is still much smaller than the zero-pressure balloons now in use). In December, Ball and his team will fly another of the same size from Antarctica in hope of breaking the 42-day duration record. In 2010, if all goes well, a balloon three times bigger will be flown, one capable of lifting an instrument payload weighing a tonne, and some small payloads will be tested. Routine flights could start in 2011.

The balloon programme office has requested that NASA officials start a dedicated line of funding for scientists seeking to build balloon-borne instruments, because some competitions for astrophysics funding at present only consider satellite missions. Cathey has no doubt that there will be an appetite. “This is the \$2.99 16-ounce steak that nobody knows about,” he says. Today's test brings that steak a good step further towards the table.

That calls for celebration: but not today. At the end of July about 40 CBSF employees will head to Shreveport, Louisiana, for three days of horse racing, gambling and golf. “This is when we will celebrate,” says Ball. “Not on campaigns. We're steely eyed balloon guys in the field.” After the adrenaline rush of the morning, Cobble heads off alone to Fred's Restaurant for a hamburger before retiring to the Billy the Kid motel. He deserves a nap.

**Eric Hand covers the physical sciences for Nature from the Washington DC office.**

G. GARDE

# CORRESPONDENCE

## Reverse translation: clearing a path from bedside to bench

SIR — Your News Feature ‘The Full Cycle’ (*Nature* **453**, 843–845; 2008), about taking research from bedside to bench, focused on unexpected results from clinical trials that stimulated a basic scientific investigation into what went wrong in those trials. Although this is a fascinating application of ‘reverse translation’, I fear it misses the much bigger opportunities and challenges that arise from researchers trying to understand clinical problems that are nowhere near the clinical-trial stage. By listening to the problems experienced by our clinical colleagues, and having the resources to address those problems in novel and creative ways, researchers can not only contribute to patient benefit, but can also introduce new ways of thinking or whole new paradigms to basic research.

But there is a sting in the tail to this approach, mentioned in your News Feature ‘Crossing the Valley of Death’ (*Nature* **453**, 840–842; 2008). Attempting to resolve apparently intractable clinical problems can be a valley of death for researchers who currently need to produce high-impact papers, reviewed mostly by basic scientists. The time taken to get to grips with a new problem, sometimes even a completely new approach, together with the necessarily less clear-cut experimental design inherent in clinical research, can become a ‘do or die’ endeavour for a biomedical scientist.

I believe the solution to this dilemma is procedural and cultural. Grant-review committees need to give a clear steer on the trade-off between many forms of clinical problem-solving and conventional criteria for scientific excellence. To this end, leaders of funding bodies must pay more than lip service to ensuring that both forward and reverse translation are placed at the

centre of their funding strategy. Longer-term funding schemes, such as the intramural support provided by many public and private funding bodies on both sides of the Atlantic, can play a specific role in promoting risky but potentially fruitful translational research. In the end, though, all parties must somehow be persuaded to sign up to both health and research.

**David R. Moore** MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, and National Biomedical Research Unit in Hearing, Nottingham University Hospitals Trust, Nottingham NG7 2UH, UK

## Translational research: don’t neglect basic science

SIR — Recent editions of *Nature* have trumpeted the merits of translational research (*Nature* **453**, 823, 830–831, 839–849; 2008) and then — almost as an afterthought to redress the balance — of basic research (*Nature* **453**, 1144, 1150–1151; 2008). You highlight concerns that increased investment in translational research could be eroding support for basic research, evidence for which is writ loud on the UK Medical Research Council’s web pages, for example. Of the many recent targeted calls for proposals, almost all are translational. As pointed out in *Nature*, this doesn’t necessarily mean less money for basic research, but it does mean that basic science misses out on a share of the substantial amount of new money.

For investigators of basic science, this trend creates a conundrum. Go where the money is, or hope that basic science retains its financial foothold. Some adroitly plant a foot in each camp — although for many this is an unsatisfactory ploy, because basic and translational research have different challenges and rewards.

Basic science is typically hypothesis-driven and leads to

new discoveries, whereas translational research applies those discoveries to patient benefits. The link between the two is exemplified by the discovery of small interfering RNA, which emerged from basic research that was eventually awarded the Nobel prize, and is now being tested as a therapeutic tool. The basic scientist may get as far as filing a patent application, but often loses interest in product development and commercialization.

Basic researchers should not have to worry about the validity of their work, but they will often add a rider to ‘sell’ their papers and grant applications — “these studies on [insert name of cell line or obscure gene] should pave the way for new treatments for [insert disease *du jour*]”.

Such concerns underline the need to include champions of basic science among science policy-makers and the bodies that fund biomedical research. Otherwise the fear is not that basic science will be deliberately run down, but that it will indeed become an afterthought and die by neglect.

**Stephen Moss** Division of Cell Biology, Institute of Ophthalmology, University College London, 11–43 Bath Street, London EC1V 9EL, UK

## Health science: from bench to bedside to trench and back

SIR — The belated push to accelerate clinical discovery through translational research is an uncertain undertaking, as you remark in your Editorial ‘To thwart disease, apply now’ (*Nature* **453**, 823; 2008).

In addition to the conversion of laboratory knowledge into new products and the adoption of such products by providers into routine clinical practice, clinical translational research must encompass behavioural and communications science in order

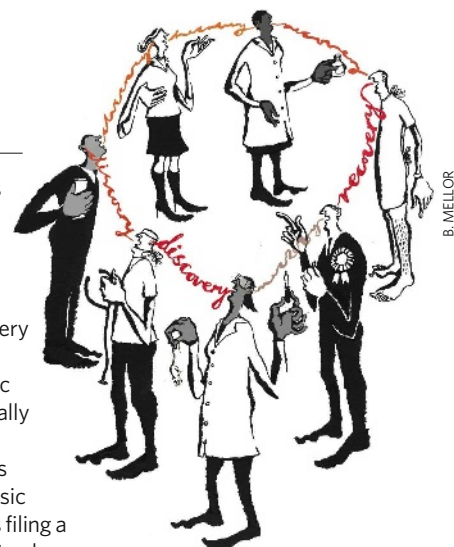
to study population uptake of new medical and scientific advances. This is crucial in a world of increasingly empowered, but not necessarily discerning, consumer patients suffering from information overload.

Community-based participatory research aims to engage the public, with a view to facilitating clinical-trial accrual and mobilizing the local expertise of community leaders. This can help to shape the investigations themselves, so that new science is both medically efficacious and effective in the real world of diverse people and patients.

The US National Institutes of Health (NIH) attempts to steer institutional culture and organizational ethics in this direction. It is supporting assessment of the effects of clinical translational science on social processes and infrastructure in academic medical centres, as well as the design of the research studies themselves.

Translational research should demonstrate a return on investment for basic science that is supported by the tax payer, and be a game-changing commitment by the NIH. The effort will need to transcend disciplinary silos to create a genuine delivery continuum for health science, from bench to bedside to trench and back again.

**Simon J. Craddock** Lee Department of Clinical Sciences, Division of Research Ethics, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9070, USA



B. MELLOR



## BOOKS &amp; ARTS

# Missing links in food-chain story

Our actions have driven top predators from much of the world, resulting in complex consequences for many ecosystems, explains **Stuart Pimm**.

**Where the Wild Things Were:  
Life, Death and Ecological Wreckage of  
Vanishing Predators**

by William Stolzenburg

Bloomsbury: 2008. 240 pp. \$24.99.

Walking in the dark across the forest clearing to my hut after dinner, I hear a tiger roar. The sound resonates through the trees making the animal sound frighteningly large, like some science-fiction monster. A herd of gaur — sleek, muscular and menacing wild cattle — then emerge from the shadows. “They, and not you, are tiger food,” tiger expert Ullas Karanth tells me, but the hairs on the back of my neck defy his reassurance. Being where predators can eat you is a rare experience, and getting rarer. What happens when they vanish from ecosystems is the subject of this entertaining book, *Where the Wild Things Were*.

Journalist William Stolzenburg describes how the removal of top predators affects species at lower levels in the food chain. He does so through the experiences of the researchers who discovered these relationships. He portrays prominent ecologists such as Robert Paine and Charles Elton in their graduate-student days, when their results were new and controversial.

*Where the Wild Things Were* starts with Paine’s 1966 influential paper on removing the starfish *Pisaster ochraceus* from rocky shores. His experiment was simple and effective, and has been imitated many hundreds of times. Paine showed that without the top predator, intertidal communities soon hold very different sets of species. Without predators, predator-sensitive species can take over communities in which they would otherwise barely survive.

On a vastly greater scale, scientists generally accept that humans exterminated many large animals shortly after their first contact with them. In the 1960s, geoscientist Paul Martin lectured on how the dung of giant sloths — “sloth shit” as he called it with relish — stopped accumulating across widely scattered caves in North and South America at much the same time, some 11,000 years ago. Martin’s result defied the conventional wisdom at the time, that noble savages could not have killed off the sloths so quickly and

over such a large area. This story was a pure historical reconstruction, for human impacts cannot be tested experimentally. However, the progression of human contact, first in Australia and later on smaller oceanic islands, provides some replication of this process.

At a scale between the neat experiments of Paine and Martin, Michael Soulé studied the absence of predators in Californian canyons accidentally isolated by the rapid growth of tract housing in San Diego. The longer the



**Star turn:** removing predatory starfish from shores allows different species to take hold.

isolation period and the smaller the canyon, the less likely were the few remaining large-bodied predators to survive. In their absence, smaller predators ran amok and eliminated ground-nesting birds. Ecologist John Terborgh investigated islands isolated by the flooding of Lake Guri in Venezuela. Similarly, a lack of predators on these islands precipitated an intricate set of cascading changes, involving plants, insects and vertebrates. Jim Estes, a marine ecologist, carefully pieced together the interactions of killer whales, sea otters, urchins and kelp by combining historical and geographical sources. Again, the message is complexity — urchins eat kelp, sea otters reduce urchin numbers and are thus good for kelp — and so on, up a long food chain involving these and other species.

Stolzenburg’s emphasis on history is also a weakness of the book, which stops short of present theory and practice. In 1960, Nelson Hairston, Fred Smith and Larry Slobodkin proposed that the world is green because predators keep herbivores under control and allow plants to flourish. Two decades later, Lauri Oksanen and his colleagues posited that a green world makes little sense if you are standing in tundras, deserts or grasslands. Where plant production is very low, the few transient herbivores eat little overall so their removal would not be noticed. In slightly more productive systems, more abundant plants support effective herbivores that can strip the greenery but are not themselves eaten by larger predators in significant numbers. Only in the most productive ecosystems would predators control herbivores and greenness prevail.

Whatever the merits of these arguments, the ubiquity of top-down predation is still being debated. Stolzenburg acknowledges this in his epilogue, but quickly concludes that predators are in control, pounding home the point with examples in every chapter. Decades of experiments show that nature is more nuanced.

Stolzenburg is forthright about the consequences. “Ecological wreckage in a land of vanishing predators” is part of the subtitle, and the back cover page notes “chaos in their absence, brazen mobs of deer, marauding raccoons, urchin-scoured reefs, [and] bizarre impoverished landscapes of pest and plague” as examples.

The loss of predators is a bad thing, but the outcomes of experiments, following Paine’s intertidal work, are more complex than this simple Armageddon. Remove a predator and the odds are nearly equal that a given prey species will flourish or flounder. A prey species can easily decrease if the predator controls another prey species that, when unchecked, can outcompete its opposition. As Bruce Menge showed in 1995 in his insufficiently appreciated synthesis, this is but one common pattern in many possibilities. Moreover, communities with more species display richer patterns.

What might we do about the loss of predators? In *Where the Wild Things Were*, Stolzenburg discusses only “re-wilding” — a fanciful attempt to return North American

S. MAFFORD/GETTY IMAGES

ecosystems to how they were before humans intervened. Natural ecological processes could be revived in US prairies, for instance, by replacing long-extinct mammoths, sabre-tooth cats and rhinos with their twenty-first-century ecological equivalents from Africa. One should not deny a man his fantasies, but the reality of restoring large species provides more immediate, yet still important challenges, including how much land is needed for various species.

The American Prairie Foundation is restocking prairies with pure-bred bison and smaller species, such as prairie dogs and their highly endangered predators, the black-footed ferret. The Wildlands Project is a US effort to connect existing wild areas from Alaska to Mexico to make them large enough to support large viable populations of predators. Its South African equivalent, the Peace Parks Foundation, notes that even the national parks of Africa are not always large enough to support lion and wild dog populations, and wants to connect them into 'mega parks'.

Another South African experience is not widely appreciated. During that nation's decades of political and academic isolation, game biologists conducted hundreds of mostly successful reintroductions of many herbivore species, including rare black rhino, black wildebeest and bontebok antelope, to their historical range in dozens of provincial parks. Numerous private game reserves have now extended those experiments, raising practical but difficult questions about how large reserves must be to hold predators or even their prey.

A few years ago, I sat munching on my lunchtime sandwich, watching wildlife that ought only appear together in a dream. Fifty wild horses stared at me. The red deer grazing behind them were oblivious. Farther away a large flock of barnacle geese had stopped to breed, and overhead soared a breeding pair of sea eagles. Wild horses have long disappeared from Europe; in the twentieth century, barnacle geese bred only in the high Arctic, and deer and eagles were found only at the sparsely populated fringes of the continent. Some 5 metres below sea level, the Oostvaardersplassen nature reserve in the Netherlands is an unlikely place to have restored something close to Europe's post-Pleistocene fauna. That these big, wild things now live there shows that nearly anything is possible. Surely, the exciting story is where the wild things were, and will be once again. ■

**Stuart Pimm** is professor of conservation ecology at the Nicholas School of the Environment, Duke University, Durham, North Carolina 27708, USA.

## Science wars revisited

**Beyond the Hoax: Science, Philosophy, and Culture**

by Alan Sokal

Oxford University Press: 2008. 488 pp.  
\$39.95, £20.00

What a difference a decade makes. In the mid-1990s, scientists and academics studying scientific culture were at each other's throats. The scientists thought the sociologists, historians and literary critics were ludicrously ignorant of science, making all kinds of nonsensical pronouncements. The other side dismissed these charges as naive, ill informed and self-serving. The exchanges became known as the science wars.

In 1996, physicist Alan Sokal landed in the centre of this fray by fooling the editors of the journal *Social Text* into publishing as a serious contribution his hilarious parody of cultural studies of science. His new collection of essays, *Beyond the Hoax*, lets us know what he has been up to on the cultural front in the decade since.

At the peak of the science wars, in 1997 Sokal and I both attended an extraordinarily interdisciplinary symposium in Santa Cruz, California. Sparks flew and proclamations of the imminent end of civilization were voiced by many in a large auditorium packed with partisans. Later that year I joined a smaller, less contentious gathering at the University of Southampton, UK, which resulted in a book. In *The One Culture? A Conversation About Science*, a diverse group, including Sokal and me, stated positions, commented on the positions of the others, and commented on the comments. I date the return of peace to academia to 2001, the year this book came out.

Other things happened in 2001 to take the steam out of the science wars. The new administration of George W. Bush decreed strict constraints on federal support for US research on embryonic stem cells, in an early example of its readiness to place ideology over science (and other forms of expertise). This provided a tough new standard against which to measure threats to science. And on a single morning, 3,000 people were murdered by terrorists in the United States in a horrifying demonstration of the real fragility of civilization. As a sign that the science wars are over, I cite the 2008 election of Bruno Latour — one of Sokal's favourite bêtes noires — to Foreign Honorary Membership in that bastion of the establishment, the American Academy of Arts and Sciences.

*Beyond the Hoax* gives us a memento of those fraught but innocent days of the 1990s. It begins with a reprint of the famous parody, accompanied by a rambling commentary that could itself be a parody of pedantic literary explication. This facing-page exegesis is set in so tiny a font that it gets farther and farther ahead of the text, ending halfway through the article so that all pages from 50 to 90 that would have had even numbers are blank. The commentary explains the jokes, teaches bits of physics, expands on cited texts, and addresses subsequent criticisms. Hoax fans ought to enjoy it.

In a preface, Sokal announces his "visceral distaste for books that have been confected by pasting together a collection of loosely connected, previously published essays". His book, he explains, is different. These ten essays (seven previously published) "form, I believe, a coherent whole". But virtually everyone who publishes a collection of essays believes they form a coherent whole. Sokal's obliviousness to this is an early indication of a complacency about his own views, and a lack of imagination about what others might be thinking, that undermines much of what follows.

Take, for example, what he says about Arkady Plotnitsky's interpretation of an obscure reply by the controversial, charismatic, deconstructionist philosopher Jacques Derrida, lampooned as meaningless nonsense in Sokal's hoax and by earlier science warriors. Derrida was asked whether Einstein's view of space-time might contain an example of a subtle Derridean concept, 'the centre of structure'.

Sokal acknowledges that Plotnitsky "has a fair knowledge of physics", but this fails to capture the unique role Plotnitsky played in the 1990s as the sole participant in the conversation who was as comfortable with theoretical physics and mathematics as he was with literary theory, sociology and science history. Plotnitsky took several pages to elucidate the technical concept of a 'centre', on which the much-maligned comment hinges, before suggesting what Derrida might have been getting at. This demonstration that Derrida's remark need not sound absurd if you are as well acquainted with Derrida as you are with Einstein, is dismissed by Sokal for three reasons: Plotnitsky gives two possible readings, "he offers no evidence that Derrida intended (or even understood) either of them", and Derrida was alive at the time so "why not just ask him?"

Or take Sokal's remarks about the physicist, biologist, and historian and philosopher of

**"Proclamations of the imminent end of civilization were voiced by many."**



science, Evelyn Fox Keller. Keller suggested there may be a style of doing science that made it more congenial to male than female practitioners. Sokal acknowledges her excellent credentials as a scientist: “[Keller] is a different can of worms entirely.” But he concludes from her reference to “the values articulated by early modern scientists” that she “implies that the issue is one of ethics, rather than of epistemology”, missing the point that epistemology is not always so easily disentangled from values. He calls her association of science with the mastery and control of nature “a gross exaggeration” on the grounds that early Newtonian mechanics applied to astronomical phenomena “over which we have no control whatsoever”. He rejects Keller’s speculation that a different style of doing science might result in a different science, because it “has no meaning except insofar as one can imagine that a different ideology might have led (in an imaginary counterfactual world) to a different course of scientific development”. But Keller’s well-known biography of the biologist Barbara McClintock is, in part, an exploration of this proposition in the real factual world.

Sokal also takes on Keller’s essay “Cognitive repression in contemporary physics”, a tour de force that invokes Jean Piaget’s views of cognitive development to shed a different light on difficulties in the interpretation of quantum mechanics. In 1991, the *American Journal of Physics* declared the essay to be one of their six memorable papers of 1979. Overlooking Keller’s wit, her fresh perspective and the fact that the paper was published nearly 30 years ago, before the current burst of interest in foundations of quantum mechanics, Sokal brusquely dismisses it as “exceedingly meager”, with views that are neither new nor insightful.

Admittedly, Plotnitsky and Keller are two of Sokal’s more formidable targets. And I should note that elsewhere in the book he is just as critical of those for whom he has sympathy as he is of those he finds — and I often agree — irredeemably foolish. He has an admirable passion for clarity of thought, and is commendably opposed to those who would pass off nonsense as profundity, whether they be commentators on science, spokesmen for religion, or the governments of nations. But Sokal’s unwillingness to expand his frame of reference to accommodate legitimately different points of view undermines his effectiveness as a scourge of genuine rubbish.

I would like to think that we are not only beyond Sokal’s hoax, but beyond the science wars themselves. This book might be a small step backwards. ■

**N. David Mermin** is Horace White Professor of Physics Emeritus at Cornell University, Ithaca, New York 14853-2501, USA. His latest book is *Quantum Computer Science: An Introduction*.



Three women in this family had breast cancer. Would they have been helped by a genetic diagnosis?

## To know or not to know?

**Blood Matters: From Inherited Illness to Designer Babies, How the World and I Found Ourselves in the Future of the Gene**  
by Masha Gessen

Harcourt: 2008. 321 pp. \$25.00

In 2004, journalist Masha Gessen published a personal account in the online magazine *Slate* of an increasingly common modern medical dilemma: how to respond to genetic diagnoses that offer both too much and not enough information. Gessen, a 37-year-old with a mutant form of the *BRCA1* gene that is associated with an increased risk of breast cancer, promised to announce in the next issue whether she would undergo preventative surgery to remove her breasts and her ovaries to reduce her chances of developing cancer. Her book *Blood Matters*, now published four years on, explores the difficulties this decision involved.

Gessen encountered similar issues to those described in an earlier first-person narrative of genetic diagnosis, *Mapping Fate* by Alice Wexler (University of California Press, 1995). Wexler chronicled her experience of living in a family affected by Huntington’s disease in the wake of the 1983 discovery of a genetic marker for the condition. She had a front-row seat from which to produce an account of what it means to have one’s fate mapped genetically: her mother died from Huntington’s, her sister is one of the leading scientists who helped to find the Huntington’s gene, and the research was partly funded by the Hereditary Disease Foundation established by their

father. At the centre of Wexler’s narrative lies a powerful lesson about the burdens of genetic knowledge. Paradoxically, the difference between abstract genetic information and real personal experience could be described as a knowledge gap. For the Wexlers, this gap was illustrated by the fact that, having found the gene and developed the test, neither sister was prepared to take it.

“To know or not to know” has thus become an added dimension of inherited pathology, creating new kinds of bonds and conflicts over the meanings of inherited genetic substance, among family members and unrelated people who share the same condition. This ambivalence is created by the very means intended to alleviate the burden of suffering to begin with — the production of detailed and accurate knowledge of how genes work. As in prenatal testing, the surfeit of explicit, readily accessible and sophisticated technical and scientific knowledge can make decisions that are based on genetic information more, rather than less, difficult.

Gessen’s exploration of her own genetic inheritance, and her world tour of genetic research facilities and hospital clinics, highlight the dilemma of genetic decision-making more vividly than any such narrative to date. She comes to realize that the frontier of genetic medicine for her condition is “essentially surgical”. She moves beyond weighing up the pros and cons to investigating the paradigms that shape genetic research, taking the reader on a breakneck tour of medical, biological, statistical, psychological and economic perspectives on genes, genetic risk

G. ROBERTS/REX FEATURES



and genetic disease. These snapshots produce a very contemporary form of autobiography — the autogenography, perhaps.

What she learns affects her ideas about herself and her children. From an apartment window in Moscow, she watches her adopted 9-year-old son drinking beer with a stranger in a park and reflects on the “dangerous business” of applying theories of genetic drift to humans when so much of how we are shaped is environmental. Is her son drawn to alcohol consumption from an early age by imitation or by genetic orientation? Like many commentators on the new genetics, Gessen is struck by the proliferation of genetic explanations in everyday life. She coins the term

‘biobabble’, after psychobabble, to describe how the gene is overused as a justification, such as when “midcareer professionals start explaining to potential partners that they had never formed an intimate relationship because of this or that polymorphism”.

Despite its futuristic subtitle, and the novelty of increasing ‘genetic choice’ as a result of developments in gene testing, the somewhat unexpected lesson from Gessen’s book is that the difficulties associated with genetic futures are not so new. They are being solved with the same combination of heart and mind that is one of the defining characteristics of being human, she finds. An unfortunate absence in Gessen’s account is the scholarly contribution

from social science or social psychology, such as in Theresa Marteau’s and Martin Richards’s landmark book *The Troubled Helix* (Cambridge University Press, 1999), which also illustrates this point. Here, and in the wide literature providing qualitative analysis of the genetic knowledge gap that separates abstract scientific facts from actual human experience, Gessen would find that the active and engaged community she describes in *Blood Matters* is even larger than she thinks. ■

**Sarah Franklin** is associate director of the BIOS Centre at the London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. She is author of *Born and Made: An Ethnography of Preimplantation Genetic Diagnosis*.



Body basics: Terry Whitlatch's fantasy creatures are inspired by the anatomy of real animals.

T. WHITLATCH

## Science artists draw together

### Annual Meeting of the Guild of Natural Science Illustrators

Ithaca College, Ithaca, New York  
20–26 July. See [www.gnsi.org](http://www.gnsi.org)

Science illustrators are in demand. Artists create images for books, posters and museum displays. In the film industry, talented concept artists design animals and fantasy creatures.

Forty years ago, the first meeting of the Guild of Natural Science Illustrators at the Smithsonian Institution, Washington DC, drew only a handful of local participants. The guild

now has almost 1,000 members and, next week, hundreds will converge on Ithaca, New York, for the 2008 conference. Through workshops, lectures and field trips, “traditional artists using pencil, crowquill pens, and watercolour will come together with those on the edge of the latest computer software doing reconstructions and 3D animations”, says Gretchen Halpert, the guild’s president.

Terry Whitlatch, a speaker at this year’s conference, used her zoology training to design the monkeys and zebras for the 1995 fantasy film *Jumanji*. Her initial images were based on realistic skeletal, muscle and surface anatomy

features that were then transformed into the three-dimensional computer-generated images seen in the film.

Whitlatch went on to invent alien creatures for the 1999 film *Star Wars Episode I: The Phantom Menace*. “I took ideas from animals and incorporated them into the world of fantasy art,” she says. “I make sure the anatomy of a creature is matched with its environment to make it believable.” The character of Jar Jar Binks, which was almost entirely computer generated for the film, was based on Whitlatch’s combination of the duck-billed dinosaur, or hadrosaur, a frog, a parrot fish for skin patterns and markings, and an emu for the character’s walk and stance.

Whitlatch creates her art with traditional techniques, then scans the artwork and enhances it



with computer software. “The computer can’t do it all for you,” she says. “It’s still just a tool like pencil and paper.”

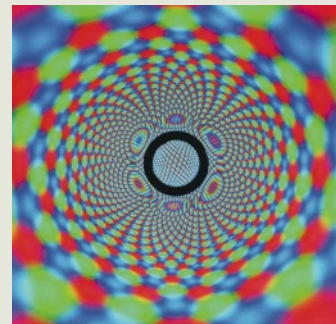
Keynote speaker Warren D. Allmon is not an artist, but an invertebrate palaeontologist. Scientific illustrators are “especially important in palaeontology, which is a mixture of science, inspired guesses, and artistic creativity to form a picture of what animals may have once looked like”. The images can also help kids to think about how scientists do palaeontology, he explains. Reconstructing ancient worlds will be the focus of another keynote speaker, James Gurney, author and illustrator of *Dinotopia: A Land Apart From Time* (HarperCollins, 1992) and the artist behind the World of Dinosaurs series of US postage stamps.

Science illustrators claim that the origin

of their craft dates back to prehistoric cave drawings. Medieval physicians, alchemists, naturalists and early scientists — including Newton and Galileo — used etchings and engravings to supplement their manuscripts. One of the best examples is Robert Hooke’s *Micrographia*. Published in 1665, it introduced readers to the microscopic world through its detailed drawings.

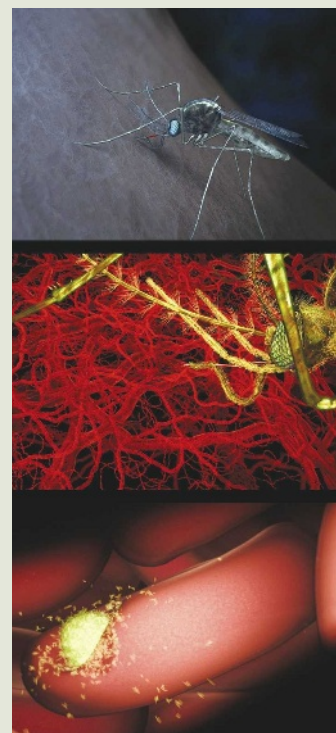
“Historically, most illustrators have worked in isolation,” adds Halpert. “The formation of the guild helped us learn what others were doing. What many think of as disparate vocations [science and art] have always been two sides of the coin.”

**Nick Thomas** is an associate professor of chemistry at Auburn University Montgomery, Alabama 36124-4023, USA.



#### AND THE WINNER IS...

The winners of the Niche Prize, a competition held by the Royal Institution of Great Britain in partnership with *Nature*, have been announced. The winning works will be displayed in two wall niches on the ground floor of the institution in London, which reopens to the public in August. *The Malaria Lifecycle* (detail, below), by Australian visual designer Drew Berry, is a scientifically accurate animation of the malaria parasite invading a human host. *Faraday’s Magnetoscope* (above) by Ken Skeldon, a research fellow at the University of Glasgow, UK, allows viewers to generate colour patterns by passing a magnet near the screen of a cathode-ray-style computer monitor, celebrating Michael Faraday’s discovery of electromagnetic induction at the Royal Institution some 180 years ago.



## Climate comedy falls flat

### Sizzle

Premieres 19 July at Outfest Film Festival, California; screens on 26 July at Woods Hole Film Festival, Massachusetts. Showing at US universities thereafter.

After watching Al Gore’s straight-faced presentation in the 2006 film *An Inconvenient Truth*, director Randy Olson turned to humour to tackle the issue of climate change. “I really liked it,” he says of Gore’s documentary in the opening of *Sizzle*, Olson’s new film that premieres in Hollywood on 19 July, “but I kept thinking ‘where are all the scientists?’”

Part documentary, part mockumentary, *Sizzle* follows Olson as he interviews scientists and sceptics from around the United States with the help of a crew of thinly drawn comedic characters. Director of the well-received 2006 documentary on intelligent design, *Flock of Dodos*, Olson is no stranger to portraying society’s complex response to science. Sadly, *Sizzle*’s mix of styles confuses his message.

The interviewees, including researchers such as Gerald Meehl at the National Center for Atmospheric Research in Colorado and Richard Somerville at the Scripps Institution of Oceanography in California, look as if they think they are being interviewed for a straight documentary. Then Olson’s cameraman, a climate-change sceptic played by comedian Alex Thomas in a solid and funny performance, butts in with his own contribution, a running gag that the film turns into a point about scientific communication: when you ask dumb questions in everyday language, scientists

suddenly start talking like ‘normal’ people.

Olson argues that data alone fail to convince people to care about climate change. What succeeds is anecdote and emotion. Indeed, the film contains very little data. The interviews are edited down to mere stubs, with no explanation of what climate change is or why we should laugh at the sceptics. Perhaps Olson assumes his audience has previous knowledge.

Olson heads to New Orleans in search of compelling anecdotes. With the caveat that the relationship between climate change and hurricanes is still being worked out, he argues like many before him, including Gore, that Hurricane Katrina is a window into a climate-changed world — even the richest nation on Earth disintegrates into anarchy in the face of terrible natural disasters. But once he gets there, Olson wanders off-topic, preferring to criticize the federal response to the hurricane.

The film is good in places and provides insights into the social response to climate change. But some of the mockumentary’s jokes, meant to keep us watching, are rather stale: an elderly white woman — Olson’s mother — goes out on the town with the young black camera crew, and the gay film producers throw several colourful hissy fits.

Ultimately, one is left wondering what the film aims to do. Does it argue that climate change is real, or discuss how we might convince people that it is? At the end of the film, Olson heads off to the editing studio to make a coherent story out of his footage. If only we had got to see that version.

**Emma Marris** is a correspondent for *Nature*.

# ESSAY

## 30 years: from IVF to stem cells

**Ruth Deech**, former chair of Britain's Human Fertilisation and Embryology Authority, reflects on how the science that gave an infertile couple a baby has been extended to saving lives.



When Louise Brown was born of the first successful *in vitro* fertilization (IVF) in 1978, the reaction was intense and mixed. Some 5,000 couples immediately signed up for the same treatment, while the Catholic Church promptly objected, drawing battle lines that are still evident today. Reporters besieged the hospital in which the birth took place. A bomb scare, perhaps called in by reporters hoping to get a glimpse of the mother, Lesley Brown, cleared the maternity building. More IVF births followed quickly, in Australia in 1980 and the United States in 1981, and a feeling arose among everyone involved that action should be taken to ensure future advances in reproductive technology would be controlled and monitored.

The British response to this has proved a massive success. By creating a climate in which embryo research is regulated and supported, Britain encouraged the use of that research to save lives.

It began in the 1980s with a move typical of how the British government handles complex issues: the establishment of a committee of the 'great and the good' charged with considering the future. The committee's 1984 report was the foundation of the full-scale legislation put in place in 1990: the Human Fertilisation and Embryology Act. This brought into being the now-famous regulatory authority the Human Fertilisation and Embryology Authority (HFEA) — a body of around 20 people, half scientists and half lay-people. The act itself could be interpreted broadly, so the HFEA had the flexibility and authority to make decisions on a case-by-case basis as reproductive medicine advanced. As I discovered while chairing the HFEA from 1994 to 2002 — one of the greatest privileges of my career — it is salutary for an ethics body to have to make immediate decisions involving real individuals and their health.

Britain is now admired internationally for its policies and practice in reproductive biology. Perhaps its main distinction is in having created a set of nationally uniform and



Louise Brown's birth in 1978 set the ball rolling for *in vitro* fertilization.

enforceable guidelines that are applicable to both research and assisted-reproduction practice. In the United States, assisted reproduction is nearly an unregulated black market, guided by toothless 'rules' from non-regulatory bodies — fertility clinics can offer sex selection at their own discretion, for example. Meanwhile, US researchers are hampered by strict laws, including the inability to use federal funds to create new embryonic stem-cell lines. The reports from the US national council on bioethics are no substitute for a regulatory body such as the HFEA, which combines reflection with statutory powers of inspection and enforcement.

### Legal sanction

The 30-year anniversary of IVF comes at the same time as a fresh bill in Britain, expected to be passed into law later this year. The bill gives legal sanction to mixed animal-human embryos, preimplantation genetic diagnosis

and saviour siblings, while banning most cases of sex selection. In practice this will change little, but it is a welcome statement that the British public and lawmakers recognize the tremendous advances made in alleviating infertility and disease by our clinicians and embryologists, and that progress in this field will remain as controlled and thoughtful as it has been in the past. It is an achievement of which Britain should be proud.

Things have not been easy for the HFEA along the way. It has faced many fights with the law and the popular press; it has struggled to keep clinicians in line, avoid commercialism, safeguard the health of donors and patients, and above all explain to the public at large the issues involved.

Ethical quandaries were raised by the science of embryo diagnosis, for example, which developed in tandem with IVF. Preimplantation genetic diagnosis meant that a couple aware of a genetic disease within the family could choose an embryo found to be clear of the threatened disease. The technique also made possible the selection of an embryo that would, if it became a baby, provide a sick older sibling with matching tissue that might save his or her life.

This was tested first in the United States with the case of Molly Nash, who was born in 1994 with Fanconi's anaemia. The only cure seemed to be a transfusion of bone marrow — but no matching donor could be found. After a long struggle, against both bureaucracy and biology, in 2000 her parents eventually conceived another baby through IVF, selected to be free of anaemia and a good tissue match for Molly.

This expensive and rarely successful procedure continues in the United States without regulation, despite concerns that it involves the disposal of dozens of 'unsuitable' embryos and opens the door to selection for other traits, from athletic ability to blue eyes. In Britain, those same concerns were raised, but with a regulatory body to oversee each decision. In 2002, the HFEA gave permission for a saviour

JOHN FROST NEWSPAPERS/ASSOCIATE NEWSPAPERS



sibling in the case of the Hashmi family, whose son had  $\beta$ -thalassaemia (sadly, their efforts to find a disease-free matching embryo met with failure). The legality of testing for saviour siblings was challenged by pro-life groups in the British courts from 2003 to 2005. But in this case as in others, the court took a decision that favoured the extension of science; it agreed that the HFEA should be free to use its discretion in allowing those advances.

The promotion of research was also the driving force behind the HFEA's decisions on animal-hybrid embryos. The HFEA has allowed (and soon British law will rubber-stamp) the insertion of human DNA into an animal egg, to address the shortage of human eggs for research. Surprisingly, this gave rise to the most public concern in the 2008 bill discussions, although most scientists saw no ethical difficulty with creating hybrid embryos that would be allowed to develop for no longer than 14 days.

### Enter the clones

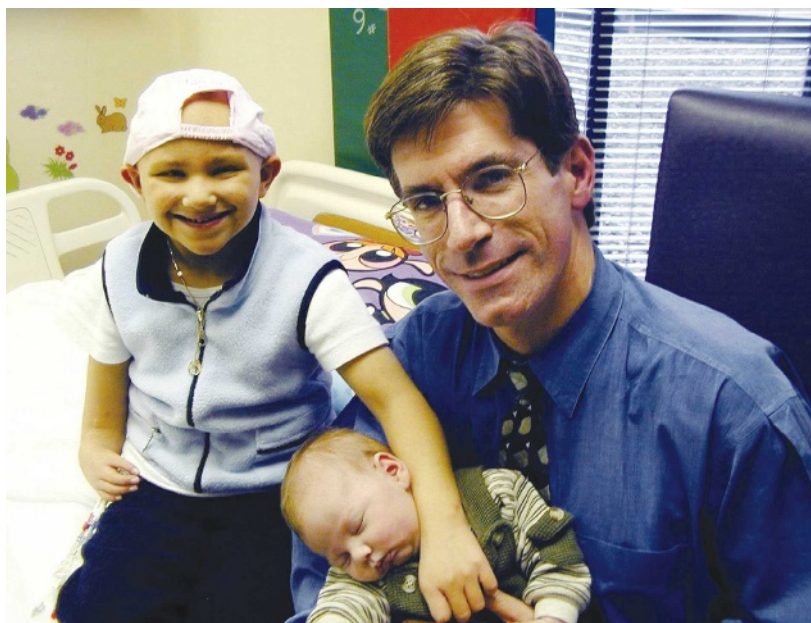
Perhaps the greatest success of the HFEA came after the cloning of Dolly the sheep in 1996 — another British 'first'. The word 'cloning' aroused considerable public concern, as the public imagined identical humans being pumped out of test tubes: an attitude engendered by films and novels. Britain aimed to embrace therapeutic opportunities arising from this success, and to rescue the word 'cloning' from the general disapprobation of 'reproductive cloning'.

After a thorough investigation of the issues by the many interested British medical and ethical organizations (including the HFEA), parliament was persuaded in 2001 to broaden the law and permit 'cell nuclear replacement' for investigating serious disease. As a result, embryo work moved from being directed solely towards infertility to searching for cures from embryo-derived stem cells. Thus the saving of life was added to the other ethical principles applicable to embryology: the safety and welfare of the child and mother, the autonomy of those involved and respect for the embryo. Britain became a safe and desirable location for stem-cell research.

In retrospect it was a clear advantage for Britain to have enacted legislation at the outset, before anyone knew what complexities lay

ahead. The act and the responsive authority were well equipped to deal with each issue as it arose. And thanks to these steps, data have been collected nearly from the start. As a result Britain has built up an unprecedented database of donors and recipients of IVF, the results of which will be of immense use to epidemiologists following the ongoing health of these children and the success of IVF techniques.

Australia also started to regulate reproductive technologies early, and has likewise been able to respond nimbly to changing situations. Some other countries have joined the legislative bandwagon a little late, largely emulating the current British system; Canada is one example.



Molly Nash (left) received healthy tissue from her IVF baby brother Adam.

In many other countries, a lack of early controls gave disagreements over controversial issues a chance to become entrenched, making it harder to formulate research-permissive — or even any — regulatory legislation. Italy, for example, went from being totally unregulated to having prohibitions based on the firm religious beliefs of the majority of the population: in 2004, embryo freezing and research were banned and IVF was restricted to heterosexual couples.

Some evidence suggests that more comprehensive national regulation, established early, accords with swift and confident technological advances. But this doesn't always hold true; the international responses have been wide-ranging. Israel and Belgium have very advanced embryology research but disjointed laws on fertility treatments. These laws are so liberal that there has been some disquiet about reproductive tourism to Belgium and

excessive treatment in Israel, which allows women almost unlimited attempts at IVF, and has the highest per-capita rate of the technique's usage. China has started to implement permissive national guidelines; the Chinese attitude towards the embryo is not burdened with Christian views.

### A long road ahead

Some US states, including California, are pushing ahead with locally regulated stem-cell work. The California Institute for Regenerative Medicine in San Francisco has an Independent Citizens' Oversight Committee, echoing the British approach. The United States as a whole,

however, has a mountain to climb if the next administration approves federal funding for stem-cell research. It has no nationwide prohibitions on keeping embryos for more than 14 days, no databases of donors and treatments, no uniform safety standards and no control on the sale and advertising of gametes. Stem-cell research can hardly be monitored in a vacuum. I predict that there will be a rush to basic registration and monitoring once the current federal ban is lifted. It will be hard to form rules that might encumber what is now a big business in assisted reproduction; rules that would differ from state to state and that will probably

attract constitutional challenges as infringing on the liberty of citizens and physicians.

Extensive legal regulation has drawbacks: expense, bureaucracy and a struggle for domination between politicians, churchmen, scientists, clinicians and the public. Yet I feel that its existence in Britain put this small country at the forefront of research and gave it an authoritative voice in international discussions on embryology. Regulation has enabled progress to be made in tandem with public acceptance and in a safe zone for clinicians and scientists who follow the rules. The passage of the 2008 bill will clear the path for scientific advances of the future: both foreseen and unforeseen. ■

**Ruth Deech** is an independent member of the House of Lords and co-author of *From IVF to Immortality: Controversy in the Era of Reproductive Technology*.

See Editorial, page 253, and News, page 260.

M. ENGEBRETSON/AFP/GETTY IMAGES

## NEWS &amp; VIEWS

## DEVELOPMENTAL BIOLOGY

## Serpent clocks tick faster

Freek J. Vonk and Michael K. Richardson

**Snakes have graceful, elongated bodies containing hundreds of vertebrae. This extreme of morphology stems from evolutionary changes in a developmental clock that regulates body patterning.**

Evolution has produced an astonishing diversity of body shapes adapted to different lifestyles. An important contributor to the shape of animals with backbones is the number of bones (vertebrae) that make up this structure. Some animals have gone to extremes (Fig. 1), none more so than snakes, which have more vertebrae than any other living animal — often more than 300, with some species<sup>1</sup> having more than 500. On page 335 of this issue, Gomez *et al.*<sup>2</sup> describe their investigations into how snakes develop this spectacular number of vertebrae. Their data suggest that a developmental 'clock', which regulates key steps in body patterning, ticks faster (relative to growth rate) in snakes than in shorter-bodied animals. This is a dramatic example of heterochrony<sup>3</sup>, in which adjustments in developmental timing lead to morphological change.

At first glance, a snake looks like a long tail with a head at one end; but snakes are not 'all tail'. Their tail — defined as the part of the body that follows the cloaca (the opening at the end of the genital and intestinal tracts) — is in fact relatively short. In front of the tail, the snake has a neck, thorax and abdomen like any other reptile, although the boundaries between these regions are not obvious. Loss of these boundaries in the snake body, along with its loss of limbs, is thought to be due to evolutionary changes in the workings of the *Hox* developmental genes<sup>4</sup>. How snakes make such large numbers of vertebrae is a separate question.

Vertebrae develop from segments of tissue called somites, which form, one after another, in a head-to-tail sequence in the embryo (Fig. 2a). They bud off from the 'head' end of the presomitic mesoderm (PSM), an immature tissue fated to generate the somites. This budding is regulated by a 'clock-and-wavefront' model, which was first proposed by Cooke and Zeeman<sup>5</sup> in 1976. Cyclical waves of gene expression, controlled by a molecular oscillator called the segmentation clock<sup>6</sup>, spread through



**Figure 1 | Serpentine silhouettes.** Body elongation, including increased numbers of vertebrae and a longer backbone, has evolved independently many times among vertebrates. **a**, Eastern glass lizard (*Ophisaurus ventralis*). **b**, European eel (*Anguilla anguilla*). **c**, Sao Tome caecilian (*Schistometopum thomense*). **d**, Malayan pit viper (*Calloselasma rhodostoma*). Photos not to scale. (**a,b,d**, F.J.V. & M.K.R.; **c**, S. Blair Hedges.)

the PSM from tail to head. These oscillations are halted when they meet a wave of maturation that sweeps through the PSM in the opposite direction, from head to tail. Each cycle results in a new pair of somites, the process ceasing when the wave of maturation catches up with the 'tail' end of the PSM.

Gomez and colleagues<sup>2</sup> compared corn snake embryos with those of shorter-bodied animals (mice, chicken and zebrafish), which have significantly fewer vertebrae, to determine whether the large number of vertebrae in snakes is generated by a variation on the clock-and-wavefront model or by some fundamentally different mechanism. They found the same molecular toolkit for body segmentation in corn snake embryos as in other vertebrates: a cyclic pattern of clock-gene expression in the developing PSM. However, subtle variation in this common theme for body segmentation in vertebrates produced the greater

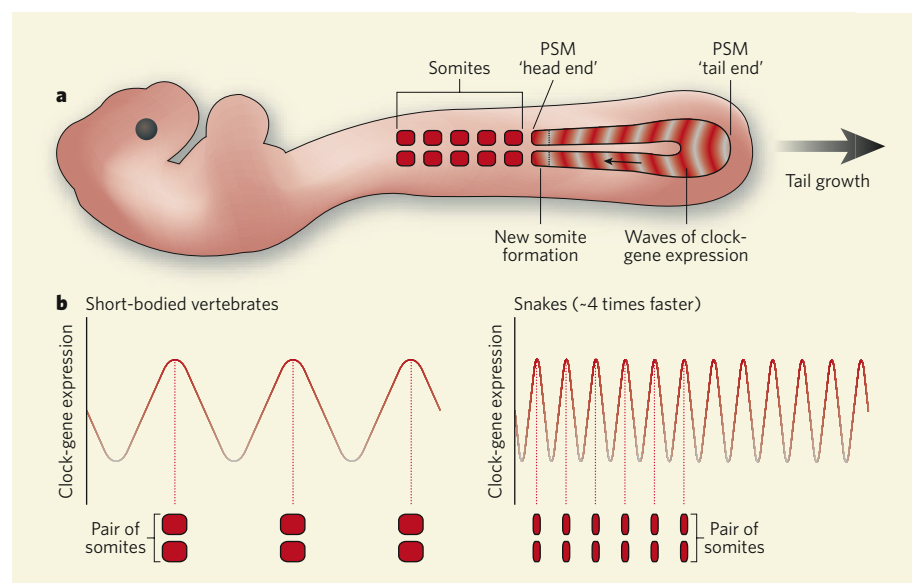
numbers of vertebrae in the snakes.

The authors found that the snake PSM develops more slowly and has a longer lifespan than in short-bodied animals. Despite this longer lifespan, the snake PSM does not undergo many more cell divisions to make the backbone. There are roughly 21 rounds of cell division in snakes, 16 in chicks and 13 in mice. This is too small a variation to fully explain the high vertebral count in snakes. But in addition, the clock component of the clock-and-wavefront mechanism is accelerated — ticking around four times faster, relative to growth rate<sup>2</sup>. Snake embryos thus segment their available backbone tissue more rapidly than other animals, forming smaller and more numerous somites (Fig. 2b).

The clock-and-wavefront mechanism may be a 'hotspot' for natural selection, because changes in vertebral number can greatly affect fitness. The number of vertebrae an animal has can influence such factors as locomotor speed and even fecundity (giving more room in a body for eggs<sup>7</sup>), and may underlie the evolution of some species groups. Serpentine body shapes have evolved independently in groups such as eels, caecilian amphibians and some lizards, as well as in snakes (Fig. 1). In these long-bodied animals, there would have been an adaptive advantage in evolving an extreme body shape, perhaps allowing different ecological niches to be conquered.

Snakes put their long bodies to good use when moving through dense vegetation, along tree branches or in water, which they do with incredible grace. Constricting snakes use their coils to suffocate prey, whereas venomous snakes are like a coiled spring that can strike prey at a distance and with terrifying speed. Snakes are among the most successful of vertebrate groups, both in terms of number of species and geographic distribution<sup>8</sup>. Elongated bodies have certainly contributed to this success, and the fact that heterochrony may be involved is particularly topical given the





**Figure 2 | Speeding-up somitogenesis.** **a**, Somites, the regular embryonic body segments from which some bones (including vertebrae), muscles and other tissues develop in vertebrates, form by a general 'clock-and-wavefront' model<sup>5</sup>. Waves of cyclical gene expression, driven by 'clock' genes, mature at the 'head' end of the presomitic mesoderm (PSM) to form paired blocks of spatially distinct somitic tissues. **b**, In snakes (right), the clock components seem to tick around four times faster (relative to growth rate) than in shorter-bodied animals (left), leading to many more, though smaller, somites<sup>2</sup>.

current interest in chronomics, the study of time regulation in biological systems.

Model species (such as mice) will always be the mainstay of research because scientists can engineer changes in their genomes and look for effects on the body plan under controlled conditions<sup>9</sup>. Biologists are nevertheless increasingly considering non-model organisms. In snakes, natural selection has done

the genetic engineering for us, and so we can study them as experiments performed by evolution. Investigating snakes hand-in-hand with model species will provide a holistic view of evolution and development. Snake embryos are not easy to work with in the lab because it is difficult to open their eggs without damaging the embryo, and obtaining eggs from snakes is not always easy. But if these technical

challenges are overcome, and snake genome sequences become widely available, a new era of 'evo-devo' research may open up.

The question remains whether the somite clock also ticks faster in other long-bodied animals, or whether natural selection has exploited alternative ways of creating a serpentine body. For example, we know that abnormal incubation temperatures can influence vertebral count in embryos (reviewed in ref. 10), and it would be interesting to know how this relates to the ticking of the somite clock. Even in snakes, alternative strategies for body elongation are possible: real giants, such as anacondas (*Eunectes murinus*), which can reach lengths of 8 metres or more, do not have more vertebrae than other snakes, they just have larger ones<sup>11</sup>. Further comparative studies in non-model species may show us how natural selection has tinkered with developmental mechanisms and pathways to produce different body plans. ■

Freek J. Vonk and Michael K. Richardson are at the Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, the Netherlands. e-mails: f.j.vonk@biology.leidenuniv.nl; m.k.richardson@biology.leidenuniv.nl

1. Marx, H. & Rabb, G. B. *Fieldiana Zool.* **63**, 1–321 (1972).
2. Gomez, C. et al. *Nature* **454**, 335–339 (2008).
3. Richardson, M. K. *BioEssays* **21**, 604–613 (1999).
4. Cohn, M. J. & Tickle, C. *Nature* **399**, 474–479 (1999).
5. Cooke, J. & Zeeman, E. C. *J. Theor. Biol.* **58**, 455–476 (1976).
6. Palmeirim, I. et al. *Cell* **91**, 639–648 (1997).
7. Shine, R. J. *Evol. Biol.* **13**, 455–465 (2000).
8. Greene, H. W. *Snakes: Evolution of Mystery in Nature* (Univ. California Press, 1992).
9. Hérault, Y. et al. *Nature Genet.* **20**, 381–384 (1998).
10. Fowler, J. A. Q. *Rev. Biol.* **45**, 148–164 (1970).
11. Head, J. J. & Polly, D. *Biol. Lett.* **3**, 296–298 (2007).

## MICROSCOPY

# Spot the atom

John Silcox

**Heavy atoms can be detected by electron microscopy, but lighter atoms, such as carbon or hydrogen, are more elusive. These bashful atoms can now be pinpointed if they are adsorbed to a single layer of graphite.**

Visualizing individual atoms is hard, not least because free atoms move rapidly at room temperature. The problem is finding a way to keep them motionless for long enough to observe them, without the resulting image being swamped by images of surrounding atoms. Ideally, they would be made to sit still on a transparent substrate in a near-perfect vacuum, so that the resulting image would be an array of sharp spots, rather like bright stars in a cloudless night sky.

Reporting in this issue (page 319), Meyer et al.<sup>1</sup> bring that ideal closer to reality. They have used electron microscopy to observe carbon atoms sitting on a graphene sheet (a single layer of carbon atoms arranged in

a 'honeycomb' array). Even more impressively, they have used image recording and computer simulations to identify hydrogen atoms — the smallest and most difficult atoms to spot — adsorbed to the sheet.

There are several reasons why graphene is superb as a backdrop for Meyer and colleagues' microscope images. First, under the conditions used for electron microscopy it traps molecules or atoms that land on its surface, so providing a way of immobilizing the specimens for long enough to get a decent image. Second, a single atomic layer is the thinnest, most transparent support possible, so any clouding of the image by the substrate is kept to a minimum. But despite being extremely thin, graphene

is also strong — certainly robust enough to survive three hours or more of irradiation by high-energy electrons, which are sometimes required to spot small atoms.

Meyer et al.<sup>1</sup> used an electron microscope that cannot identify carbon atoms in a graphene lattice directly, but detects the change in intensity that results when an atom binds to the graphene. The difference in intensity is caused by the extra scattering of incident electrons by the adsorbed atom, which adds to the number of electrons scattered by the graphene layer itself. In the authors' images, the carbon atoms show up as dark spots against a bright background — a negative image of a night sky (see Fig. 2a on page 320).

Meyer and colleagues' technique allows real-time observation of atomic and molecular dynamics, and of chemical reactions that occur under electron irradiation. The authors demonstrate this by recording a series of images showing the formation of vacancies in the graphene sheet — holes in the lattice caused by missing atoms — as the carbon monolayer is bombarded with electrons. They even record one of these vacancies disappearing as it is filled by a carbon atom adsorbed on the graphene sheet. No doubt

other processes, involving the interactions of various adsorbed atoms with each other and with graphene, could be observed directly with this technique. Also, the use of atomic sheets other than graphene will widen the range of reactions that can be studied. A method for the routine preparation of diverse two-dimensional crystals is still some way off, but several different kinds of sheet have been prepared<sup>2</sup>.

The authors recorded videos showing the interaction of linear, carbon-based molecules (thought to be hydrocarbons) with the incoming electron beam. Their achievement opens up the possibility of using electron microscopy to study molecular dynamics, although this would have its limitations. For example, because graphene is two-dimensional, the number of configurations adopted by adsorbed molecules might be limited compared with the number that occur on a three-dimensional substrate or in a vacuum.

There are other ways to visualize atoms, but researchers using such techniques could learn a trick or two from Meyer and colleagues' approach. One particularly promising method<sup>3</sup> determines the locations of atoms in big molecules by analysing the diffraction patterns of electrons that are scattered by those molecules. Unfortunately, this works only for samples that are suspended in a vacuum. Extending this diffraction method<sup>3</sup>, so that it could analyse molecules adsorbed to a simple, well-understood substrate (as in Meyer and colleagues' strategy<sup>1</sup>), would allow a much wider range of molecules to be visualized at the atomic scale.

Because of graphene's remarkable resistance to damage by electron bombardment, structural features can be recorded many times over. The resulting images can then be superimposed, creating a sharper picture than could be obtained from a single image. Certain atoms bond strongly enough to graphene to stick around for similarly long periods, and so can also be treated in this way. In such cases, excellent data can be obtained and compared with calculated predictions of electron scattering, to confirm the identity of the atom doing the scattering. Indeed, one compelling application of Meyer and colleagues' work would be to obtain accurate information on how electrons are scattered by different atoms. Such data are sorely needed for assessing the accuracy of theoretical models, and for allowing a more accurate identification of atoms by electron-scattering methods. There is a slight possibility, for example, that the hydrogen atoms detected by the authors could actually be helium atoms. By collecting more accurate data, any uncertainty about this could be ruled out.

Occasionally, a paper is published that opens up lines of investigation previously thought impossible; this is just such a paper. As the authors comment<sup>1</sup>, electron microscopy has benefited tremendously from recent advances in instrumentation, but the elephant in the room is that these advances amount to nothing if you can't find the right sample to analyse. By showing what can be done using

graphene as a substrate for specimen preparation, Meyer *et al.* might have filled a gap that has limited the full potential of electron microscopy for decades.

John Silcox is in the School of Applied and Engineering Physics, Clark Hall, Cornell University, Ithaca, New York 14853, USA.

e-mail: js97@cornell.edu

1. Meyer, J. C., Girit, C. O., Crommie, M. F. & Zettl, A. *Nature* **454**, 319–322 (2008).
2. Novoselov, K. S. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).
3. Zuo, J. M., Vartanyants, I., Gao, M., Zhang, R. & Nagahara, L. A. *Science* **300**, 1419–1421 (2003).

## MOTOR NEURON DISEASE

# The curious ways of ALS

Magdalini Polymenidou and Don W. Cleveland

**That mutations in the SOD1 enzyme underlie inherited forms of a motor neuron disease known as ALS is clear. But the question of what the consequences of such mutations are seems to have more than one answer.**

Amyotrophic lateral sclerosis (ALS) is a neurological disorder characterized by the selective premature degeneration and death of motor neurons, which control voluntary actions such as breathing and walking. The disease starts in adult life, and the ensuing progressive paralysis is typically fatal within a few years, usually owing to failure of the respiratory system. Reporting in *Genes & Development*, Nishitoh *et al.*<sup>1</sup> provide new insight into the molecular events that provoke ALS, linking its pathogenesis to a well-established stress pathway in an intracellular organelle called the endoplasmic reticulum.

Although most cases of ALS are sporadic, in roughly 10% of instances the disease is dominantly inherited — that is, carrying even one copy of the disease-associated mutation is sufficient to cause this fatal paralysis. A landmark discovery<sup>2</sup> in 1993 ushered in the molecular era of ALS research, with the finding that mutations in the gene encoding the enzyme superoxide dismutase 1 (SOD1) are responsible for 20% of inherited ALS cases. (Normal SOD1 is a crucial intracellular antioxidant, facilitating the clearance of the potentially toxic superoxide radical.)

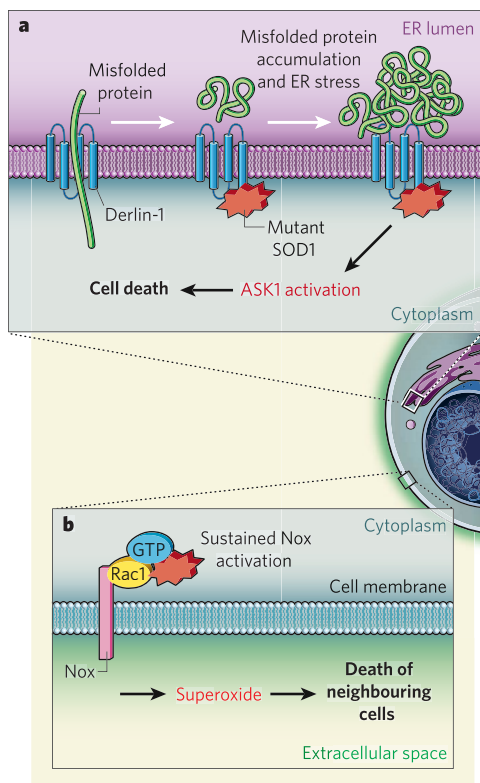
Mice expressing various ALS-related mutations in SOD1 recapitulate the fatal paralysis seen in human patients. Studies in these animals have established that ALS is caused by the toxic activity of the SOD1 mutation rather than reduced activity of this enzyme<sup>3</sup>. Selective silencing of the mutant gene in various cell types has shown that the ubiquitously expressed mutant SOD1 damages motor neurons and non-neuronal neighbouring cells such as astrocytes and microglia. Whereas damage to motor neurons is associated with the onset of ALS, damage to astrocytes and microglia — the innate immune cells of the spinal cord — sharply accelerates the progress of the disease<sup>4,5</sup>. But it is a sobering reality that, 15 years after the discovery of the SOD1 mutation in familial ALS, no consensus has emerged as to the molecules and pathways that are directly affected by mutant SOD1.

Nishitoh and colleagues<sup>1</sup> now show that the mutant enzyme inhibits ERAD, the cell's machinery for eliminating proteins that fail to fold properly in the endoplasmic reticulum (ER), following up on earlier claims that ER stress is a pathogenic component in ALS<sup>6</sup>. The first crucial step in the 'ERADication' of misfolded proteins is their transport from the lumen of the ER to the cytoplasm, where their covalent attachment to the protein ubiquitin (a process known as ubiquitination) marks them for degradation by the proteasome complex. The authors find that three different SOD1 mutants, but not the normal enzyme, can interact with an ERAD component called Derlin-1, which is instrumental in both the transport of misfolded proteins from the ER lumen to the cytoplasm and their degradation. By binding to Derlin-1, mutant SOD1 inhibits the ERAD pathway, causing both defective functioning of the ER (ER stress) and activation of a protein kinase called ASK1, which is involved in driving programmed cell death (Fig. 1a). These observations provide a putative direct molecular link between mutations in SOD1 and a main pathological hallmark of ALS — the presence of ubiquitinated aggregates of misfolded proteins in the cytoplasm.

The precise cells in which Derlin-1 and mutant SOD1 interact are not known, although the most likely candidates are astrocytes rather than motor neurons, because deletion of the ASK1 gene prolongs survival of a SOD1 mutant mouse by slowing disease progress, without affecting its onset. But a perplexing finding is that the interaction between Derlin-1 and mutant SOD1 is not detectable until symptoms of the disease appear, suggesting that it is secondary to some unidentified initiating event.

Nishitoh and colleagues' observations follow several other recent findings. Earlier this year, it was reported<sup>7</sup> that both normal and mutant SOD1 associate with Rac1, an activator of the enzyme NADPH oxidase (Nox) found at the cell membrane. Normally, Nox mediates the production of superoxide — in immune cells





**Figure 1 | Proposed mechanisms of toxicity in SOD1-mediated ALS.** **a**, Nishitoh *et al.*<sup>1</sup> show that mutant SOD1 interacts with Derlin-1, a protein that is essential for transporting misfolded proteins from the lumen of the endoplasmic reticulum (ER) to the cytoplasm for degradation. Consequently, misfolded proteins accumulate in the ER lumen, and the resulting ER stress activates ASK1 — a protein involved in programmed cell

death — by an unidentified mechanism. **b**, Near the cell membrane, mutant SOD1 also interacts strongly with Rac1 (ref. 6), keeping the Nox enzyme active. The resulting increased production of the superoxide radical, and its release into the extracellular space, is toxic to neighbouring cells. The precise cells in which mutant SOD1 interacts with Derlin-1 or Rac1 are unknown.

two neurons). Excessive stimulation of the glutamate receptors on neuronal membranes, and the accompanying increase in calcium-ion influx, can trigger a cascade of toxic events in the neuron, including damage to both mitochondria (the cell's powerhouses) and the ER (the main reservoir of intracellular calcium). Loss of the glutamate transporter EAAT2, a synaptic 'vacuum cleaner' for glutamate, from neighbouring astrocytes contributes to this phenomenon<sup>8</sup>, but a direct role — if any — for mutant SOD1 in this mechanism remains unclear.

Mutant SOD1 has also been proposed<sup>9</sup> to interact directly with chromogranins (components of vesicles secreted by neurons) and to co-secrete with them. This in turn may trigger damage to motor neurons and neighbouring cells.

And there is more. Mitochondrial dysfunction has long been implicated in the damage to motor neurons seen in ALS. Misfolded

mutant SOD1 is deposited on the cytoplasmic face of the outer mitochondrial membrane in cells of the spinal cord<sup>10</sup>. Moreover, mitochondrial dysfunction in motor neurons<sup>11</sup> carrying SOD1 mutations is associated with the release of cytochrome c, a potent trigger of cell death. Finally, mutant SOD1 provokes oxidative stress in astrocyte mitochondria<sup>12</sup>, which could accelerate death in neighbouring neurons.

With all these potential contributors, the proposed pathways of pathogenesis in inherited ALS might strike some as "curiouser and curiouser", much as Alice proclaimed of the strange happenings in Lewis Carroll's *Alice's Adventures in Wonderland*. Among the biggest challenges will be to distinguish mechanisms that cause the disease from those that are only epiphenomena. In our view, the most likely possibility is that every set of observations discussed here is partly right, with age-dependent, selective toxicity requiring a convergence of damage to motor neurons and non-neurons alike. So stay tuned: the main pieces of this puzzle are yet to be revealed. ■

Magdalini Polymenidou and Don W. Cleveland are at the Ludwig Institute for Cancer Research, and the Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-00670, USA.  
e-mail: dcleveland@ucsd.edu

1. Nishitoh, H. *et al.* *Genes Dev.* **22**, 1451–1464 (2008).
2. Rosen, D. R. *et al.* *Nature* **362**, 59–62 (1993).
3. Bruijn, L. I. *et al.* *Science* **281**, 1851–1854 (1998).
4. Boillée, S. *et al.* *Science* **312**, 1389–1392 (2006).
5. Yamanaka, K. *et al.* *Nature Neurosci.* **11**, 251–253 (2008).
6. Kikuchi, H. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 6025–6030 (2006).
7. Harraz, M. M. *et al.* *J. Clin. Invest.* **118**, 659–670 (2008).
8. Rothstein, J. D., Van Kammen, M., Levey, A. I., Martin, L. J. & Kuncel, R. W. *Ann. Neurol.* **38**, 73–84 (1995).
9. Urushitani, M. *et al.* *Nature Neurosci.* **9**, 108–118 (2006).
10. Vande Velde, C., Miller, T. M., Cashman, N. R. & Cleveland, D. W. *Proc. Natl Acad. Sci. USA* **105**, 4022–4027 (2008).
11. Pehar, M. *et al.* *J. Neurosci.* **27**, 7777–7785 (2007).
12. Cassina, P. *et al.* *J. Neurosci.* **28**, 4115–4122 (2008).

such as microglia — to kill bacteria and other pathogens. It emerges that the association of normal SOD1 with Rac1 is part of a tightly regulated mechanism that, under chemically reducing conditions, activates Nox. But mutant SOD1 interacts with Rac1 with a higher affinity than normal, 'locking' Nox in its active, superoxide-producing form, even under oxidizing conditions. This results in a tenfold increase in superoxide production and its release into the extracellular space (Fig. 1b).

Paradoxically, therefore, instead of carrying out its normal job of removing intracellular superoxide, mutant SOD1 might increase extracellular levels of this radical and so damage motor neurons and other cells. A Nox inhibitor improved survival in a SOD1 mutant mouse almost exclusively by delaying the onset of ALS<sup>7</sup>. As the presence of mutant SOD1 in motor neurons, but not microglia, is one of the factors that determine the timing of disease onset, this finding suggests that the interaction between mutant SOD1 and Rac1 triggers disease by activating a Nox variant (Nox1) found in motor neurons, rather than another variant of this enzyme (Nox2) that occurs in microglia.

So, are ER stress<sup>1,6</sup> and the increased production of superoxide<sup>7</sup> — both potential triggers of death for motor neurons and neighbouring cells — the whole story of the molecular events underlying ALS? Almost certainly not. Observations in SOD1 mutant mice and in tissue samples from patients with sporadic ALS suggest that one contributory factor to the damage is excitotoxicity (excessive firing of motor neurons that occurs when the stimulatory neurotransmitter glutamate is not rapidly removed from the synaptic junctions between

## EARTH SCIENCE

# Volcanic cause of catastrophe

Timothy J. Bralower

**From the timing, it looks as if an episode of marked oceanic oxygen deficiency during the Cretaceous was the result of undersea volcanism. Studies of such events are relevant to the warming world of today.**

About 93 million years ago, Earth was shaken by an immense episode of volcanism. Massive piles of highly fluid lava accumulated under the seabed, forming much of the present-day Caribbean region in a geological heartbeat. On page 323 of this issue<sup>1</sup>, Turgeon and Creaser argue persuasively that these eruptions triggered an episode of oceanwide anoxia that led to mass extinctions. Moreover, at this time Earth was so hot that palm trees grew in Alaska

and large reptiles roamed northern Canada. So this new work can provide valuable lessons about the way Earth responds to perturbations akin to those it is experiencing now.

The stack of lava flows that formed the Caribbean tectonic plate is known as a large igneous province (LIP)<sup>2,3</sup>. Other LIPs altered Earth's history by causing profound changes in the composition of the atmosphere: by increasing carbon dioxide levels and causing

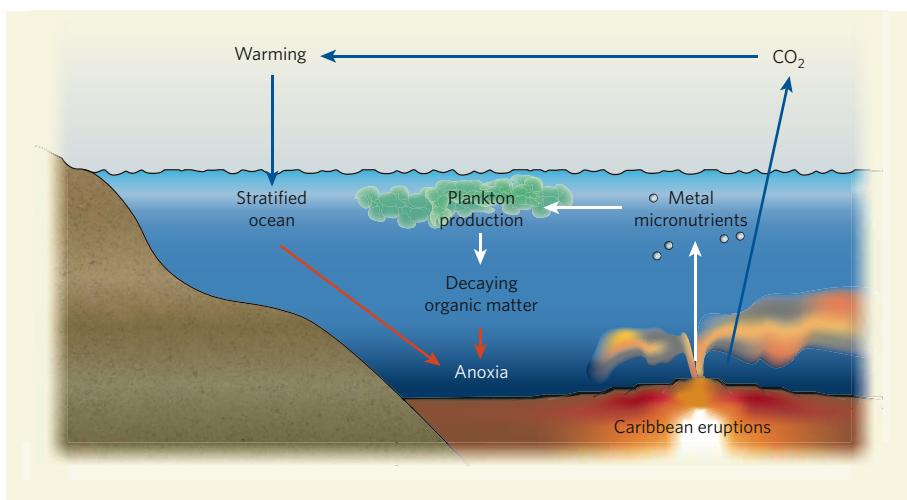
greenhouse warming; by altering the chemistry of sea water and ocean circulation; and by disrupting the global carbon cycle<sup>4</sup>. A LIP in Siberia that erupted 250 million years (Myr) ago led to the largest mass extinction ever, at the end of the Permian period<sup>5</sup>. Another event 120 Myr ago, the formation of the massive Ontong Java Plateau in the western Pacific Ocean<sup>6</sup>, led to extinctions of ocean plankton. Both of these LIPs caused the ocean circulation to slow, decelerating the cycling of carbon and oxygen, and ultimately leading to highly toxic, anoxic conditions on the sea floor.

Episodes of anoxia, known as oceanic anoxic events (OAEs), have occurred periodically during Earth's history, but none was more severe than that which occurred 93 Myr ago, during the Cretaceous period<sup>7,8</sup>. This OAE caused the extinction of large clams known as inoceramids and tiny protists called foraminifera that lived on the sea floor. Profound changes in ocean circulation also led to the production and preservation of enormous quantities of marine organic material that was subsequently transformed into oil during its burial. But the ultimate cause of the OAE has proved elusive. There is strong evidence that warming was involved, implicating greenhouse conditions<sup>9</sup>. Unusual metal enrichments in rocks deposited during the OAE suggest a link to volcanism<sup>10,11</sup>, but rocks that are organic-rich are often metalliferous as well, so this link remains unconfirmed.

The Caribbean lavas are now deeply buried in the ocean or found in mountain belts in places such as Haiti, where they have been exhumed during tectonic activity. Their eruption history is not as well known as that of the Siberian and Ontong Java LIPs, which are more widely sampled. The lavas' ages, measured using radioactive isotopes, span the interval 87–95 Myr ago, with suggestions that a large pulse occurred 93–94 Myr ago<sup>12</sup>. But given that the Caribbean is hundreds of thousands of cubic kilometres in extent, it is impossible to determine its eruption history with only a few age estimates. Instead, geologists must rely on other chemical signals in sedimentary rocks that provide continuous proxies for volcanic intensity through time.

Turgeon and Creaser<sup>1</sup> used a sensitive proxy for volcanism based on two isotopes of osmium, <sup>187</sup>Os and <sup>188</sup>Os. The main sources of osmium to the ocean include detritus from rivers, volcanism and material from outer space such as meteorites. Thus, the osmium isotope composition of sea water is a measure of the weathering of the continents, volcanic activity and extraterrestrial input<sup>13</sup>. The abundance of both osmium isotopes in sea water is low, so their ratios adjust rapidly to changes in input.

Turgeon and Creaser's careful analysis of these isotopes in sedimentary rocks from drill cores off the coast of South America, and from mountains in Italy, provides clear evidence of a perturbation that immediately preceded the OAE 93 Myr ago. The isotope values show an



**Figure 1 | Volcanism, oceanic anoxia and global warming.** Turgeon and Creaser<sup>1</sup> provide good evidence for a causal connection between the extensive eruptions in the Caribbean region 93 million years ago and the oceanic anoxic event of that time. But how might they have been connected? One possibility is that the volcanism seeded the upper ocean with metal micronutrients, increasing phytoplankton production, which in turn led to increased oxygen use during the decay of organic matter. Another, not mutually exclusive, possibility is that a consequence of the global warming stemming from volcanically produced CO<sub>2</sub> was a more stratified ocean, in which oxygen delivery to deep waters became restricted.

unmistakable increase in the osmium contribution from a meteoritic or volcanic source. In the absence of other evidence for an extraterrestrial impact at this time, the data clearly point to a volcanic episode. And given that the shift in osmium isotopes suggests a factor of 30–50 increase in the osmium flux to the oceans, that episode was evidently on a huge scale. Although there are several candidates, the only LIP close to this age that is large enough to have caused this type of perturbation is in the Caribbean.

The formation of the Caribbean is of great interest to geologists, but it also has much broader implications. Abrupt warming events in the geological record are of great significance for scientists working on modern global warming. The ancient episodes provide a complete picture of the processes operating during various stages of a global warming event. For example, at some stage photosynthetic plankton will draw down CO<sub>2</sub> from the atmosphere, producing organic matter that will be buried in rocks, and leading to global cooling that possibly heralds the end of the warming event<sup>8,9</sup>. In this regard, the event of 93 Myr ago can serve as a test run for the possibility of seeding the modern ocean with nutrients to promote photosynthesis and CO<sub>2</sub> reduction.

But a better understanding of that event and its relevance to modern global change requires knowledge of the scale of the volcanism (and so the rate and amount of CO<sub>2</sub> input), and how it is related to the OAE. These questions are only partially addressed by Turgeon and Creaser<sup>1</sup>. They show that there is a temporal relationship between the LIP and the OAE that implies a causal connection, but they offer no proof for the nature of that connection.

One possibility is that the large amounts of metal-rich fluids produced during volcanism

may have seeded the ocean with micronutrients such as iron, stimulating plankton to produce large quantities of organic matter<sup>11</sup>. Oxidation of this organic matter would have stripped the ocean of oxygen, leading to the OAE (Fig. 1). Whether such a process alone could cause a global anoxic event has yet to be determined, but I remain doubtful that it could. Perhaps that effect was exacerbated by oceanic stratification (Fig. 1), a result of the warming produced by volcanic CO<sub>2</sub> that severely inhibited mixing in the ocean.

Especially significant is the time lag between the initial climatic perturbation, produced by the emission of huge quantities of CO<sub>2</sub> into the atmosphere, and its long-term consequences. Based on the rate at which the sedimentary rocks are known to have accumulated, Turgeon and Creaser<sup>1</sup> estimate that the volcanic pulse preceded the OAE by 23,000 years. Although the lag between trigger and catastrophe might be heartening to those concerned about the impacts of global warming, the time it takes signals to mix in the oceans today (about 1,500 years) implies that the lag time was shorter; other abrupt warming events in the geological record have similar, shorter lags. If the response time to the LIP volcanism was indeed longer, it suggests either that ocean circulation was extraordinarily sluggish before the event, or that there was a complicated connection between the trigger and the OAE. Determining how the volcanism of 93 Myr ago wreaked havoc on the warm Earth is another challenge for geologists endeavouring to understand the complex behaviour of our planet.

Timothy J. Bralower is in the Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. e-mail: bralower@psu.edu



1. Turgeon, S. C. & Creaser, R. A. *Nature* **454**, 323–326 (2008).
2. Coffin, M. F. & Eldholm, O. *Rev. Geophys.* **32**, 1–36 (1994).
3. Kerr, A. C. in *Treatise on Geochemistry* Vol. 3 (ed. Rudnick, R. L.) 537–565 (Elsevier, 2003).
4. Larson, R. L. *Geology* **19**, 547–550 (1991).
5. Bowring, S. A. *et al.* *Science* **280**, 1039–1045 (1998).
6. Tarduno, J. A. *et al.* *Science* **254**, 399–403 (1991).
7. Jenkyns, H. C. *J. Geol. Soc. Lond.* **137**, 171–188 (1980).
8. Arthur, M. A., Dean, W. E. & Pratt, L. M. *Nature* **335**, 714–717 (1988).
9. Forster, A., Schouten, S., Moriya, K., Wilson, P. A. & Sinninghe Damsté, J. S. *Paleoceanography* **22**, PA1219, doi:10.1029/2006PA001349 (2007).
10. Orth, C. J. *et al.* *Earth Planet. Sci. Lett.* **117**, 189–204 (1993).
11. Snow, L. J., Duncan, R. A. & Bralower, T. J. *Paleoceanography* **20**, PA3005, doi:10.1029/2004PA001093 (2005).
12. Sinton, C. W. & Duncan, R. A. *Econ. Geol.* **92**, 836–842 (1997).
13. Peucker-Ehrenbrink, B. & Ravizza, G. *Terra Nova* **12**, 205–219 (2000).

## SIGNAL TRANSDUCTION

# Linking nutrients to growth

Vittoria Zinzalla and Michael N. Hall

**How cells sense nutrients to control growth is largely unknown. One missing link involved in conveying the nutrient signal to the TOR protein, which regulates growth, seems to be the Rag proteins.**

In mammalian cells, nutrients (such as amino acids), growth factors and cellular energy together trigger a molecular signalling pathway, mediated by the protein TOR, that controls cell growth. Deregulation of this pathway is implicated in cancer, and TOR inhibition by the anticancer drug rapamycin prevents unruly cell growth. Extensive research has led to characterization of many components of this signalling pathway. A remaining question is how amino acids activate TOR. Two teams (Kim *et al.*<sup>1</sup> reporting in *Nature Cell Biology* and Sancak *et al.*<sup>2</sup> writing in *Science*) now provide an important clue.

TOR ('target of rapamycin') is an evolutionarily highly conserved protein kinase that is found in two functionally and structurally distinct multiprotein complexes: TORC1 and TORC2 (ref. 3). TORC1 controls many cellular processes that ultimately determine cell growth, including protein synthesis, ribosome formation, nutrient transport and autophagy (a survival mechanism that kicks in in response to starvation).

Activation of TORC1 requires simultaneous availability of amino acids, growth factors and energy. Although inputs from growth factors (such as insulin) and energy are relatively well understood, determining the way that amino acids trigger TORC1 activation has been elusive. Amino-acid depletion results in rapid dephosphorylation of two molecules downstream of TORC1, S6K and 4E-BP, whereas addition of amino acids leads to rapid, TORC1-dependent phosphorylation of these molecules. But what is the molecular link between the amino-acid signal and TORC1 activation?

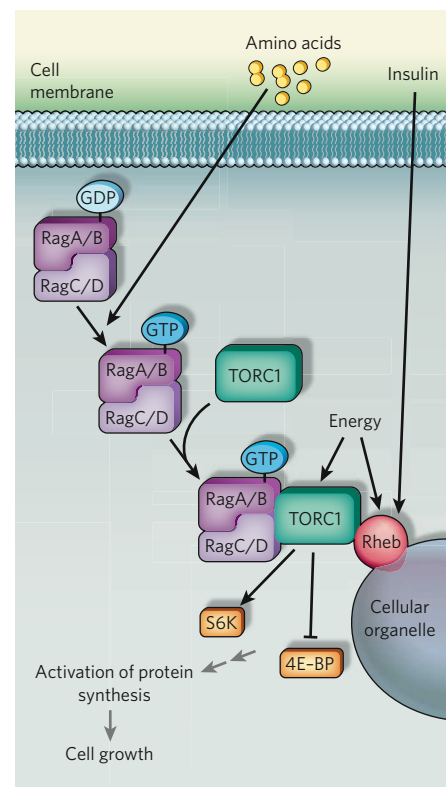
To answer this question, Kim *et al.* and Sancak *et al.* used complementary approaches. Using the technique of RNA interference (RNAi), Kim and colleagues<sup>1</sup> performed a screen with the S2 cell line of the fruitfly *Drosophila* to search for GTPase proteins that regulate S6K phosphorylation in response to amino acids. Sancak *et al.*<sup>2</sup> performed a proteomic

analysis in mammalian cells to identify new binding partners of TORC1. Both teams independently identified Rag GTPase proteins as mediators of the amino-acid signal to TORC1. Sancak and colleagues further showed that these regulatory molecules interact with rapTOR, a component of TORC1. Meanwhile, Kim *et al.* concentrated on the physiological aspects of the Rag GTPases' function, demonstrating their role in TORC1-mediated regulation of autophagy and cell size in *Drosophila*.

Regulatory GTPases come in different varieties and are commonly components of signalling pathways. These proteins are active when bound to the nucleotide GTP and inactive when bound to GDP — the product of GTP hydrolysis. The intrinsic enzymatic activity of the GTPases converts the GTP to GDP. Specific GTPase-activating proteins (GAPs) stimulate this intrinsic enzymatic activity of the GTPases, and another group of proteins called GTPase exchange factors (GEFs) mediate dissociation of GDP from GTPases so that the GTPases can bind to a new GTP molecule and resume activity. Thus, GAPs inhibit and GEFs stimulate the signalling function of GTPases.

There are four Rag proteins (RagA–D), with high sequence similarity existing between RagA and RagB, and between RagC and RagD. These proteins function as heterodimers — RagA or B binding to RagC or D (ref. 4). Rag heterodimers form independently of GTP/GDP binding status. Nonetheless, GTP binding to the RagA/B subunit is crucial for the activation and localization of the heterodimer. The two studies<sup>1,2</sup> show that a RagA/B that cannot bind to GTP also fails to stimulate TORC1, and a constitutively active RagA/B supports TORC1 activity even in the absence of amino acids.

But what is the consequence of Rag binding to TORC1? Sancak *et al.* propose that amino-acid-activated Rags are molecular match-makers for TORC1 and another GTPase called Rheb. When bound to Rag, TORC1 is somehow delivered to Rheb, which is localized on



**Figure 1 | Passing on the message.** Amino acids are transported into the cell where they trigger a molecular signalling pathway that is mediated by the multiprotein complex TORC1. Two studies<sup>1,2</sup> show that Rag GTPases — heterodimers of RagA or RagB and RagC or RagD — are involved in TORC1 activation in response to this nutrient signal. By binding to TORC1, Rag GTPases mediate its transfer to intracellular membranes that contain another GTPase, Rheb. There, other signals such as cellular energy and growth factors (insulin) integrate with the amino-acid signal, leading to Rheb-mediated activation of TORC1 and phosphorylation of its downstream effectors S6K and 4E-BP, which ultimately lead to protein synthesis and cell growth.

the membrane of a so far ill-defined organelle. Rheb then directly stimulates the kinase activity of TORC1, which leads to S6K and 4E-BP phosphorylation. Amino acids impinge on this mechanism at the level of GTP loading of RagA/B (Fig. 1).

Previous work<sup>5</sup> has shown that the Rheb–TORC1 interaction depends on amino-acid availability, and that TORC1 inhibition after amino-acid withdrawal can be overcome through increased Rheb expression. Intriguingly, inactivation of the TSC1–TSC2 complex, a GAP that inhibits Rheb, cannot overcome the effect of amino-acid withdrawal<sup>6</sup>, suggesting that amino acids probably mediate TORC1 interaction with activated Rheb, rather than activation (GTP loading) of Rheb. (Insulin and energy mediate the activation of Rheb through inhibition of TSC1–TSC2 GAP activity.)

The molecular mechanism by which Rheb activates TORC1 once Rheb has integrated amino acid, insulin and energy signals, and has bound to TORC1, remains to be determined.

An earlier study suggested the involvement of a protein called FKBP38. In the absence of amino acids, FKBP38 binds to and inhibits TORC1, and dissociates only on amino-acid-stimulated binding of Rheb to TORC1 (ref. 7). It is difficult to imagine how FKBP38 might be involved in Rag-mediated activation of TORC1, as FKBP38 is itself membrane bound. Another protein, VPS34, also seems to mediate amino-acid stimulation of mammalian TORC1 activity<sup>8</sup>. But the relationship between VPS34 and the Rag proteins in activating mammalian TORC1 is unclear, as VPS34 is not involved in activating TORC1 in *Drosophila*<sup>9</sup>.

Of the other remaining questions, perhaps the most important is: what mediates amino acids' effect on the Rag proteins? So far, no GEF or GAP specific for regulation of the Rag family of proteins has been identified. This topic is likely to become the subject of considerable attention. The finding that Gtr1 and Gtr2, yeast proteins related to RagA/B and RagC/D, are implicated in nutrient sensing by TOR also cannot be overlooked<sup>10</sup>.

Previous work<sup>11</sup> has shown that a high-protein diet (that is, increased availability of amino acids) affects glucose metabolism and insulin signalling through the TOR signalling pathway, leading to increased blood glucose levels and insulin resistance. Unravelling regulation of the Rag GTPases will help us to gain a better understanding of metabolic disorders such as obesity and type 2 diabetes. ■

Vittoria Zinzalla and Michael N. Hall are in the Biozentrum, University of Basel, CH4056 Basel, Switzerland.  
e-mail: m.hall@unibas.ch

1. Kim, E., Goraksha-Hicks, P., Li, L., Neufeld, T. P. & Guan, K.-L. *Nature Cell Biol.* doi:10.1038/ncb1753 (2008).
2. Sancak, Y. et al. *Science* **320**, 1496–1501 (2008).
3. Wullschlegel, S., Loewith, R. & Hall, M. N. *Cell* **124**, 471–484 (2006).
4. Gao, M. & Kaiser, C. A. *Nature Cell Biol.* **8**, 657–667 (2006).
5. Long, X. et al. *J. Biol. Chem.* **280**, 23433–23436 (2005).
6. Smith, E. M. et al. *J. Biol. Chem.* **280**, 18717–18727 (2005).
7. Bai, X. et al. *Science* **318**, 977–980 (2007).
8. Gulati, P. et al. *Cell Metab.* **7**, 456–465 (2008).
9. Juhasz, G. et al. *J. Cell Biol.* **181**, 655–666 (2008).
10. Dubouloz, F. et al. *Mol. Cell* **19**, 15–26 (2005).
11. Patti, M.-E. & Kahn, B. B. *Nature Med.* **10**, 1049–1050 (2004).

## SYSTEMS BIOLOGY

# On the cell cycle and its switches

Silvia D. M. Santos and James E. Ferrell

**For the cell-division cycle to progress, hundreds of genes and proteins must be coordinately regulated. Systems-level studies of this cycle show that positive-feedback loops help to keep events in sync.**

The cell cycle is a complex but orderly sequence of events that culminates in the production of two cells from one. In eukaryotes, the cycle is divided into four phases: cell growth in G1 phase, DNA replication in S phase, more growth in G2 phase, and cell division in mitosis or M phase. The system of regulators that drives transitions between phases is centred on the cyclin-dependent kinases (CDKs), enzymes that are activated when regulatory proteins called cyclins bind to them. The CDK network directly or indirectly orchestrates coordinated regulation of proteins and genes involved in essentially every aspect of cell function. The complexity of these regulatory events raises the question of what systems-level strategies keep the process temporally coherent — how does the maestro of the cell cycle generate a definitive downbeat? Writing in this issue, Skotheim *et al.*<sup>1</sup> and Holt *et al.*<sup>2</sup> examine different phases of the cell cycle in the budding yeast *Saccharomyces cerevisiae*, and their findings converge on the same answer: positive feedback.

In budding yeast, the cell cycle begins with the synthesis of the initiator cyclin Cln3, which binds to and activates Cdk1. Substrates of the Cln3–Cdk1 complex include the SBF and MBF gene transcription factors (activated through

phosphorylation), and the transcriptional inhibitor Whi5, which is translocated out of the nucleus (and so inactivated) after phosphorylation. The reciprocal regulation of SBF/MBF and Whi5 brings about the transcription of hundreds of genes that collectively constitute the G1/S regulon.

Among the targets of SBF/MBF are the genes that encode the G1 cyclins Cln1 and Cln2. Like Cln3, these two cyclins can activate Cdk1. And like Cln3–Cdk1, Cln1/2–Cdk1 complexes can activate SBF/MBF and inhibit Whi5. Thus, the Cdk1 system has a pair of interlinked positive-feedback loops that could, in principle, function as an irreversible, bistable trigger, with Cln1 and Cln2 promoting their own accumulation in an ever-accelerating cycle (Fig. 1a). The 'explosive' kinetics of the positive-feedback system could provide the definitive downbeat that keeps the G1/S regulon coherent.

This attractive idea was tested more than a decade ago in cell populations<sup>3,4</sup> and found to be (apparently) incorrect: cells lacking the *CLN1* and *CLN2* genes (*cln1Δ cln2Δ* cells) seemed to activate the promoter sequence for *CLN2* just as quickly as normal cells. But sometimes studying individual cells can reveal things that are masked by averaging over a

population, and Skotheim *et al.*<sup>1</sup> (page 291) show that this is the case here.

Examining normal cells individually by live-cell fluorescence microscopy, these authors find that Whi5 can abruptly translocate out of the nucleus some 40–50 minutes after the start of the G1 phase, and that the *CLN2* promoter is turned on at about the same time. By contrast, in *cln1Δ cln2Δ* cells, the redistribution of Whi5 to the cytoplasm occurs slowly and gradually, and activation of the *CLN2* promoter is delayed by around 40 minutes. These observations indicate that Cln1/2-mediated positive feedback is required for Whi5 to be rapidly switched off and for timely induction of *CLN2* expression.

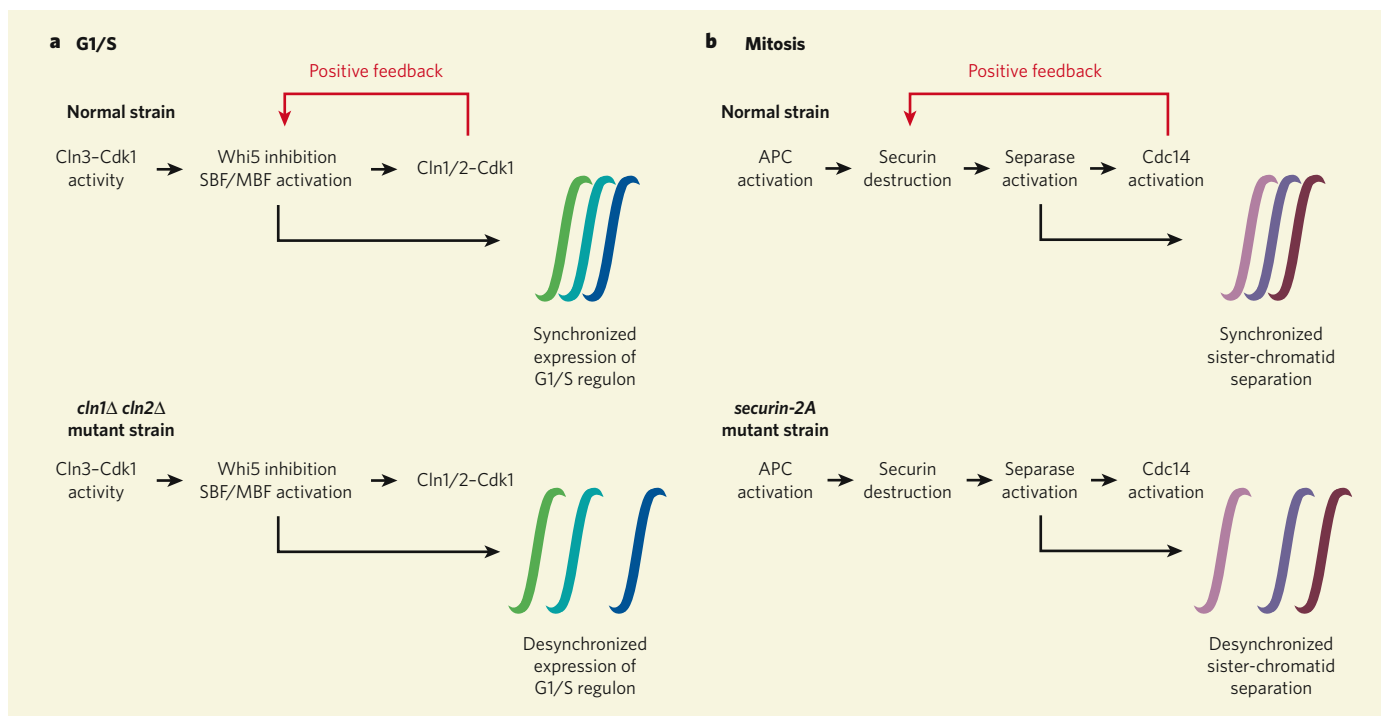
Skotheim and colleagues also find<sup>1</sup> that the expression of the G1/S regulon is desynchronized and incoherent in *cln1Δ cln2Δ* cells. Normally, the *CLN2*, *RFA1* and *RAD27* genes are turned on within a few minutes of each other. In *cln1Δ cln2Δ* cells, however, they splutter on one by one. Moreover, the most strongly incoherent yeast cells almost invariably arrest as unbudded G1-phase cells, never progressing into the S and M phases. Together, these results suggest that positive feedback makes the redistribution of Whi5 abrupt, which allows the G1/S regulon to be expressed synchronously and the cell cycle to proceed.

A similar story emerges from studies of another irreversible transition step in the cell cycle: the separation of sister chromatids at the onset of anaphase during mitosis. This event is mediated by the enzyme separase, which is normally inhibited by another protein called securin. At the onset of anaphase, the anaphase-promoting complex (APC) brings about securin's destruction, allowing separase to be activated. Sister-chromatid separation generally seems to occur with near-perfect synchrony, and Holt *et al.*<sup>2</sup> (page 353) provide an explanation for why this is so: securin is part of a previously unrecognized positive-feedback loop.

The starting point for these authors' work was the identification of a new Cdk1-mediated phosphorylation site in securin. Phosphorylation of this evolutionarily conserved site makes securin a poor destruction substrate for the APC. But once some securin has been destroyed (owing to a graded increase in APC activity), separase is activated and, in turn, activates Cdc14 — an enzyme that can dephosphorylate many CDK substrates, including securin. Securin dephosphorylation further increases the rates of its own breakdown, and of separase and Cdc14 activation. This ever-accelerating positive-feedback system could, in principle, translate a gradual increase in APC activity into an abrupt 'switch off' of securin activity (Fig. 1b).

To test this idea, Holt *et al.* tagged two chromosomal loci on chromosomes IV and V with green fluorescent protein, and followed the timing of sister-chromatid separation by fast live-cell imaging. They find that, under normal conditions, chromosome-IV





**Figure 1 | Cell-cycle synchronizers.** **a**, Skotheim *et al.*<sup>1</sup> show that positive feedback between Cln1/2–Cdk1 and the transcriptional regulators Whi5 and SBF/MBF is essential for synchronous expression of the G1/S regulon, which is disrupted when positive feedback is brought to a halt in mutant yeast strains that lack the genes *CLN1* and *CLN2* (*cln1Δ cln2Δ*). **b**, Holt and colleagues<sup>2</sup> find that positive feedback between Cdc14 and securin helps to make another step in the cell cycle — separation of sister chromatids at the end of mitosis — happen in unison. In the *securin-2A* mutant, where this feedback mechanism is lost, sister-chromatid separation does not occur as synchronously.

sisters separate from each other an average of 90 seconds before chromosome-V sisters do. But when the feedback was compromised, either through expression of *securin-2A* — a phosphorylation-site securin mutant that is resistant to positive feedback — or by manipulation of the activities of Cdc14 or Cdk1, this lag period doubled. Moreover, the rate of chromosome mis-segregation rose dramatically in the *securin-2A* strain. This is consistent with the idea that positive feedback in securin destruction is crucial for the synchronicity of sister-chromatid separation, and that perhaps this synchronicity is essential for the fidelity of genomic segregation.

These studies<sup>1,2</sup> underscore the importance of single-cell, real-time approaches for understanding the dynamics of molecular signalling networks. But insight can also be obtained from population-based approaches. Recently, Haase and co-workers<sup>5,6</sup> looked at what links G1-phase events to those later in the cell cycle. One plausible idea is that control is handed from one cyclin–Cdk1 complex to the next in temporal succession (Cln3–Cdk1 to Cln1/2–Cdk1 to the six B-type cyclin (Clb1–6)–Cdk1 complexes), ultimately resulting in APC activation, degradation of mitotic cyclins, and the resetting of the cell cycle to G1 phase. But the authors show<sup>6</sup> that, in a yeast strain lacking all six *CLB* genes, most of the cell-cycle-regulated genes are still periodically expressed. For many of the genes, the amplitudes of the oscillations are compromised, but, astonishingly, most of them still come and go with near-normal

timing. The authors propose that yeast cells have an autonomous transcriptional oscillator as well as the more familiar CDK oscillator.

So how is this transcriptional oscillator wired? Haase and colleagues<sup>6</sup> hypothesize that the circuit is a long, slow, positive-feedback loop, with *CLN1* and *CLN2* promoting their own expression through the intermediacy of a cascade of transcriptional regulators with shorter, quicker, negative-feedback loops that keep each wave of transcriptional activation pulsatile. In a sense, this is the opposite of the usual CDK-centred cell-cycle models<sup>7</sup>, which assume that a slow, negative-feedback loop (Cdk1 activation leading to APC activation leading to Cdk1 inactivation) is coupled to several short, positive-feedback loops (such as the Whi5 and securin subcircuits<sup>1,2</sup>). In support of their idea, Haase and colleagues show<sup>6</sup> that a Boolean model (a model that assumes all genes and proteins flip-flop between digital 'on' and 'off' states) of this network exhibits robust oscillations.

The positive-feedback loops examined in these studies<sup>1,2,6</sup> are not the only ones involved in cell-cycle regulation. Experiments in frog egg extracts and human cell lines have established<sup>8–10</sup> the importance of a positive-feedback loop that regulates the mitotic activation of Cdk1 in establishing stable, robust oscillations. And in yeast, the securin-mediated positive-feedback loop sits within another positive-feedback loop in which Cdk1 inactivation promotes activation of a form of the APC that feeds back to inactivate more Cdk1 molecules<sup>11</sup>. Positive feedback can help to make

transitions between states decisive and irreversible<sup>12–14</sup>, suppress chatter during transitions<sup>15</sup>, and endow oscillator circuits with robustness and tunability<sup>16</sup>. The work of Skotheim *et al.*<sup>1</sup> and Holt and colleagues<sup>2</sup> adds another performance characteristic of positive feedback to this list: it can help to keep things in sync. This powerful organizational strategy may be essential for a process as beautiful and as complicated as the cell cycle. ■

Silvia D. M. Santos and James E. Ferrell are in the Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305-5174, USA.  
e-mail: james.ferrell@stanford.edu

1. Skotheim, J. M., Di Talia, S., Siggia, E. D. & Cross, F. R. *Nature* **454**, 291–296 (2008).
2. Holt, L. J., Krutchinsky, A. N. & Morgan, D. O. *Nature* **454**, 353–357 (2008).
3. Dirick, L., Böhm, T. & Nasmyth, K. *EMBO J.* **14**, 4803–4813 (1995).
4. Stuart, D. & Wittenberg, C. *Genes Dev.* **9**, 2780–2794 (1995).
5. Haase, S. B. & Reed, S. I. *Nature* **401**, 394–397 (1999).
6. Orlando, D. A. *et al.* *Nature* **453**, 944–947 (2008).
7. Tyson, J. J., Csikasz-Nagy, A. & Novak, B. *BioEssays* **24**, 1095–1109 (2002).
8. Cross, F. R., Archambault, V., Miller, M. & Klovstad, M. *Mol. Biol. Cell* **13**, 52–70 (2002).
9. Pomerening, J. R., Kim, S. Y. & Ferrell, J. E. *Jr Cell* **122**, 565–578 (2005).
10. Pomerening, J. R., Ubersax, J. A. & Ferrell, J. E. *Jr Mol. Biol. Cell* doi:10.1091/mbc.E08-02-0172 (2008).
11. Cross, F. R. *Dev. Cell* **4**, 741–752 (2003).
12. Ferrell, J. E. Jr & Machleder, E. M. *Science* **280**, 895–898 (1998).
13. Gardner, T. S., Cantor, C. R. & Collins, J. J. *Nature* **403**, 339–342 (2000).
14. Xiong, W. & Ferrell, J. E. *Jr Nature* **426**, 460–465 (2003).
15. Thron, C. D. *Biophys. Chem.* **57**, 239–251 (1996).
16. Tsai, T. Y.-C. *et al.* *Science* **321**, 126–129 (2008).

## OBITUARY

# John Templeton (1912–2008)

Philanthropist at the interface of religion and science.

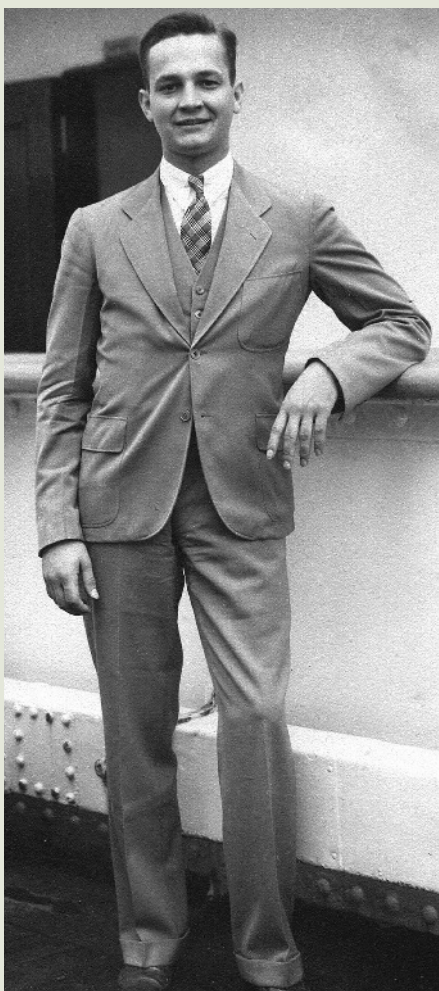
Sir John Templeton, who died on 8 July at the age of 95, was one of the most modest billionaires you could hope to meet. I interviewed him once, when he was visiting England to talk to Prince Philip — in a small stuffy room at the top of a small stuffy club on Green Park, London, where he was staying because he liked it. He could have bought the Ritz next door without noticing, but that would have been extravagant. He would have thought it wrong. He was prepared to spend billions on his causes, but never a penny more than he thought was needed to get the job done.

The job that needed doing, he thought, was to recognize the spiritual aspects of the world as quite as real, as reliable and as worth studying as anything else that science can examine. He would say that he wanted from his philanthropy “new spiritual information”. This is a belief that some atheists and many religious believers — especially those who see in modern science a conspiracy against revealed truth — find laughable or sinister. If I defend it, I will certainly be accused of an interest in the matter, as I have taken Templeton money twice, once as the first winner of his prize for European religious journalism, and once as one of his journalistic scholars of science and religion for two months. But money is actually the least that anyone invests in the question.

Whether or not you believe that the facts of the world are wholly amoral will shape everything about your life. It doesn't, of course, follow that anyone who believes the world is amoral must themselves feel compelled to amorality. You can perfectly well suppose that humans should be good in an amoral world, and you may even suppose that humans should be religious in such a world. But this will always feel arbitrary, false to the facts, not on the winning side.

Conversely, if you believe, with Templeton, that virtue is rewarded, and that a patient, honest exploration of the Universe will reveal spiritual values, you may not want to subscribe to any organized religion, but you will certainly suppose that goodness is the right and natural condition of our lives and that it is evil that demands an explanation.

That kind of optimism was one of the things the whole world loved about the United States in the 1950s and 1960s, when Templeton was making his fortune on Wall Street (\$10,000 invested with his main fund in 1954 would have grown to \$2 million when he sold out in 1992, netting him a personal profit of \$440 million; by that time his total funds were worth \$13 billion).



Templeton in 1934, en route to Oxford to take up his Rhodes scholarship.

All of this was earned from nothing. He was born in 1912 in a small town in Tennessee, 100 kilometres from Dayton, where the ‘Scopes monkey trial’, which tested the law forbidding the teaching of evolution, was held; in later life he would boast that he had known John Scopes after the trial. His father was a lawyer who lost most of his money in the Depression; the family were devout Presbyterians, as Sir John remained throughout his life. From Yale, he won a Rhodes scholarship to the University of Oxford, which seems to have given him a lifelong case of Anglophilia: in 1968 he renounced his American citizenship to become a British subject and moved to the Bahamas. He was knighted in 1987. His first wife, with whom he had three children, died in a bicycling accident in 1951; he remarried in 1958, but outlived his second wife too.

His first notable act of philanthropy was to set up an annual prize for “Progress in

Religion” which he stipulated should be bigger than the Nobel prizes, thus proving to vulgar materialists that the spiritual mattered. The first recipient was Mother Teresa of Calcutta, in 1973, to whom he gave \$85,000 (the prize is now worth well over \$1 million) — not a woman one could accuse of progress in theology but she certainly got things done, and this appealed greatly to Templeton's distinctive combination of pragmatism and energetic optimism. More recent winners have tended to be scientists with an interest in the metaphysical, among them Freeman Dyson and John Polkinghorne.

Templeton was sure that in the moral and spiritual realm there could be facts of the matter, and that science could find them. He didn't believe in a literal heaven or hell, or a Garden of Eden, but he was certain that these stories were records of encounters with aspects of reality that demand explanation. He would open the meetings of his fund with a prayer, not because he wanted God to reward them but because he believed that prayer cleared the mind and led to better decisions, even if he could not know why. He would not have been in the least bit disconcerted by the celebrated Templeton-funded experiment on the effects of intercessory prayer on a large sample of heart patients, which showed only that patients who knew they were being prayed for had slightly worse outcomes.

In 1987, he set up a foundation for research into what he liked to call the really big questions. Now with an endowment of \$1.5 billion, it has funded an enormous range of research, from cosmology to the cognitive roots of religious belief, and is credited with founding the field of forgiveness research — the effect forgiveness may have on a subject's physical health and well-being — by careful pump-priming funding. The list of moral qualities that he thought science should be looking at was entirely characteristic of the man: “ethics, love, honesty, generosity, thanksgiving, forgiving, reliability, entrepreneurship, diligence, thrift, joy, future-mindedness, beneficial purpose, creativity, curiosity, humility, and awe.”

This is slightly absurd, but at the same time noble. The world would be better off with more of all these things, and he wanted science, with his money, to study them, because he never for a moment doubted that virtues were as real as quarks. He really deserved the title of philanthropist.

**Andrew Brown**

Andrew Brown is a writer on science and religion.  
e-mail: [laptop@thewormbook.com](mailto:laptop@thewormbook.com)

See also Editorial, page 253.



# Positive feedback of G1 cyclins ensures coherent cell cycle entry

Jan M. Skotheim<sup>1</sup>, Stefano Di Talia<sup>1</sup>, Eric D. Siggia<sup>1</sup> & Frederick R. Cross<sup>2</sup>

**In budding yeast, *Saccharomyces cerevisiae*, the Start checkpoint integrates multiple internal and external signals into an all-or-none decision to enter the cell cycle. Here we show that Start behaves like a switch due to systems-level feedback in the regulatory network. In contrast to current models proposing a linear cascade of Start activation, transcriptional positive feedback of the G1 cyclins Cln1 and Cln2 induces the near-simultaneous expression of the ~200-gene G1/S regulon. Nuclear Cln2 drives coherent regulon expression, whereas cytoplasmic Cln2 drives efficient budding. Cells with the *CLN1* and *CLN2* genes deleted frequently arrest as unbudded cells, incurring a large fluctuation-induced fitness penalty due to both the lack of cytoplasmic Cln2 and insufficient G1/S regulon expression. Thus, positive-feedback-amplified expression of Cln1 and Cln2 simultaneously drives robust budding and rapid, coherent regulon expression. A similar G1/S regulatory network in mammalian cells, comprised of non-orthologous genes, suggests either conservation of regulatory architecture or convergent evolution.**

Positive feedback in genetic control networks can ensure that cells do not slip back and forth between either cell cycle phases or developmental fates. For example, commitment to sporulation in budding yeast is driven by transcriptional positive feedback of the meiotic inducer *IME1* (refs 1–3). In *Xenopus laevis*, positive feedback underlies the all-or-none characteristics of oocyte maturation<sup>4,5</sup> and mitotic entry<sup>6,7</sup>, suggesting the frequent use of positive feedback to regulate cellular transitions.

Absent from this list of examples is the well-studied Start checkpoint controlling cell cycle commitment in budding yeast. Nutrient limitation and pheromone exposure arrest cells before DNA replication, whereas size control extends G1 in small daughter cells<sup>8–11</sup>. Beyond Start, cells proceed through division almost independently of size and environment<sup>9</sup>. Previous experiments suggested that Start represents a feedback-free cascade of events<sup>12</sup> (see schematic in Fig. 1a; omitting red arrows). The transition is initiated by the G1 cyclin Cln3 (refs 13–15), which in complex with Cdc28 activates the transcription of about 200 genes<sup>16</sup> by phosphorylating promoter-bound protein complexes that include the transcription factors SBF and MBF<sup>17</sup> and the transcriptional inhibitor Whi5 (refs 18 and 19). Phosphorylation and inactivation of Whi5 is rate-limiting, and phosphorylated Whi5 rapidly exits the nucleus. The G1/S regulon, which includes two additional G1 cyclins *CLN1* and *CLN2*, contributes to the activation of B-type cyclins, DNA replication, spindle pole body duplication and bud emergence. Mitotic B-type cyclins then inactivate SBF<sup>20</sup> and, with *NRM1*, inactivate MBF<sup>21</sup>, thus turning off the G1/S regulon.

Any one of the three G1 cyclins suffices to activate the regulon, suggesting that there is potential for transcriptional positive feedback of *CLN1* and *CLN2* on their own expression<sup>22,23</sup>. However, analysis of synchronized populations led to the conclusion that positive feedback, defined as Cln1 and Cln2 advancing transcription from the *CLN2* promoter, did not occur in wild type; instead, Cln3 was the sole activator of firing<sup>14,15</sup>.

In sharp contrast to the prevailing linear model, we demonstrate that Cln1- and Cln2-dependent positive feedback is central to Start control. We use single-cell time-lapse fluorescent microscopy to

show that Cln1 and Cln2 advance timing and reduce variability in the activation of *CLN2*, and of the entire G1/S regulon. We further explore the mechanisms and functional significance of this control.

## Positive feedback of G1 cyclins

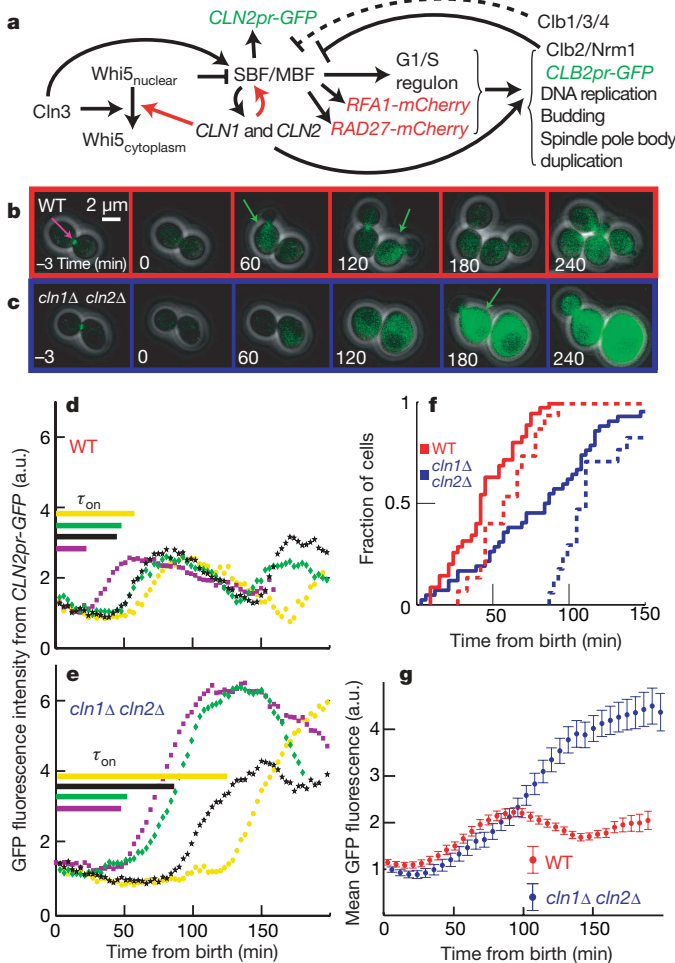
Positive feedback of Cln1 and Cln2 on their own transcription should yield faster accumulation of *CLN2* messenger RNA in wild-type cells than in *cln1Δ cln2Δ* cells. Although Cln1- and Cln2-dependent positive feedback was clearly demonstrated in the absence of Cln3 (refs 22–24), this does not indicate that wild-type cells function similarly. In synchronized populations, near-identical timing of onset of *CLN2* promoter activity was observed in the presence or absence of *CLN1* and *CLN2*, leading to the linear model<sup>14,15</sup>. Here we revisit this issue using single-cell assays. We used unstable green fluorescent protein (GFP) driven by the *CLN2* promoter (*CLN2pr-GFP*) as a reporter for *CLN2* transcription<sup>24,25</sup> (see Methods and Supplementary Figs 1 and 2). Birth time was determined using the disappearance of the Myo1-GFP myosin ring<sup>11</sup>, a marker for cytokinesis that did not influence the *CLN2pr-GFP* signal. The timing of *CLN2* promoter induction in individual cells is sharp and easily quantified computationally (see Methods, Fig. 1b–e and Supplementary Figs 1 and 2). Because *cln1Δ cln2Δ* cells are larger than wild type, we integrated *MET3pr-CLN2* in both strains to conditionally express Cln2 before time-lapse imaging so that initial sizes were comparable<sup>14</sup> (see Methods and Supplementary Figs 3 and 12 for controls). Thus, we can assay for positive feedback by comparing the time interval from birth to transcriptional activation of *CLN2pr-GFP* transcription in the first cell cycle after *MET3pr-CLN2* is turned off in wild-type and *cln1Δ cln2Δ* cells.

Positive feedback should advance *CLN2* promoter activation in wild-type compared to *cln1Δ cln2Δ* cells<sup>14,15</sup>. In daughter cells, the average time between birth and *CLN2* promoter activation ( $\tau_{on}$ ; Fig. 1d–f) was much shorter for wild type (41 min) than for *cln1Δ cln2Δ* (83 min). Furthermore, *CLN2pr-GFP* activation was much less variable for wild-type than for *cln1Δ cln2Δ* cells (standard deviation of 21 min versus 47 min). *CLN2pr-GFP* transcription was Cln3-dependent in *cln1Δ cln2Δ* cells because *cln1Δ cln2Δ cln3Δ* cells failed to induce *CLN2pr-GFP*. Qualitatively similar results were

<sup>1</sup>Center for Studies in Physics and Biology, The Rockefeller University, <sup>2</sup>The Rockefeller University, 1230 York Avenue, New York 10065, USA.

obtained in mother cells and also in cells growing in glycerol and ethanol instead of glucose. In all cases, the interval from birth to *CLN2pr-GFP* activation was smaller and less variable in wild type than in *cln1Δ cln2Δ*, indicating strong positive feedback of Cln1 and Cln2 on their own transcription independent of nutrient conditions or cell type (Supplementary Table 3;  $P < 10^{-4}$ ).

We explored the potential redundancy of *CLN1* and *CLN2* in activating the feedback loop. Although budding is slightly delayed in *cln1Δ CLN2* and *CLN1 cln2Δ* cells compared to wild type, the timing of *CLN2* promoter activation is similar (Supplementary



**Figure 1 | Positive feedback drives the Start of the budding yeast cell cycle.** **a**, Schematic of the Start transition; new interactions demonstrated in this paper are shown in red. **b**, **c**, Combined phase and fluorescence images for *CLN2pr-GFP MYO1-GFP MET3pr-CLN2* cells, either wild type (WT; **b**) or *cln1Δ cln2Δ* (**c**), grown without methionine (inducing) and plated on methionine (repressing) to normalize initial conditions<sup>6</sup> (Supplementary Fig. 3). Green arrows indicate approximate peak GFP expression from *CLN2pr-GFP*. **d**, **e**, Single-cell fluorescence intensity (in arbitrary units, a.u.) for four characteristic cells of each genotype; cells are synchronized at birth and marked by the disappearance of a Myo1-GFP ring at the bud-neck (purple arrow in **b**). The time from birth to *CLN2* promoter activation (as defined in Methods),  $\tau_{on}$ , for each individual cell is indicated by length of the corresponding line. **f**, Cumulative distribution of *CLN2pr-GFP* induction indicates that Cln1- and Cln2-dependent positive feedback contributes substantially to the early expression of *CLN2*; dashed lines indicate induction of *CLB2pr-GFP* marking the onset of negative regulation of *CLN2*. **g**, Average fluorescence intensity for 87 wild-type and 83 *cln1Δ cln2Δ* daughter cells aligned at birth simulates a population study, which would obscure the effect of positive feedback. The results shown are for daughter cells in glucose; changes in cell type or nutrient conditions do not qualitatively influence the results (Supplementary Table 3). Error bars, s.e.m.

Table 3), indicating that *CLN1* and *CLN2* form redundant conduits for positive feedback.

Our data can be reconciled with previous work<sup>14,15</sup> arguing against positive feedback because measurements averaged over a population of cells necessarily lose information. In addition to delayed onset of transcription, *cln1Δ cln2Δ* cells express a more intense and prolonged *CLN2pr-GFP* signal. The larger peaks are probably due to a delay in the Clb2-mediated repression of SBF/MBF<sup>14,15,20,21</sup> (Fig. 1d, e), because the average time between induction of *CLN2* and *CLB2* was much larger in *cln1Δ cln2Δ* strains (measured using a *CLB2pr-GFP* cassette; Fig. 1f and Supplementary Fig. 13), and Clb2 accumulation is known to be delayed in *cln1Δ cln2Δ* strains<sup>14</sup>.

Therefore, imperfect synchrony<sup>11</sup> allows the high and lengthened transcriptional response from the first *cln1Δ cln2Δ* cells firing the *CLN2* promoter to mask the delayed response of the majority. This effect is reconstituted in Fig. 1g by averaging our measured single-cell data, and explains why positive feedback was not detected in measurements of mRNA levels in populations of synchronized daughter cells<sup>14,15</sup>.

### Coherent regulon expression

Once a cell senses the signal to initiate the cell cycle, it must actuate all the machinery effecting the cell cycle transition. At Start, this requires activation of many SBF- and MBF-regulated genes<sup>16</sup> encoding proteins involved in DNA replication and bud-site formation. However, noise in protein expression in individual cells<sup>26</sup> could interfere with expression of this large regulon. In particular, the delayed and variable induction of the *CLN2* promoter in *cln1Δ cln2Δ* cells suggested that G1/S regulon expression might be severely disrupted in these feedback-free cells.

To investigate regulon expression in individual cells, we compared induction of *CLN2pr-GFP* and *RAD27-mCherry*, another member of the regulon<sup>16</sup> (Fig. 2a–d and Supplementary Figs 7 and 8). *RAD27* expression is Cln-dependent (Supplementary Fig. 11). *CLN2* and *RAD27* are synchronously induced in wild type, whereas there is a long and variable period of time between the inductions of the two genes in the *cln1Δ cln2Δ* mutant (Fig. 2e, f). Out of the 86 *cln1Δ cln2Δ* cells studied, 11 failed to produce a detectable increase in Rad27-mCherry and 4 failed to produce a detectable increase of either marker. We performed identical experiments on strains containing *CLN2pr-GFP* and *RFA1-mCherry*, another regulon member<sup>16</sup>, and obtained similar results (Fig. 2g, h). Our conclusions are valid even after excluding outlying points ( $P < 0.01$ ). Thus, Cln1- and Cln2-dependent positive feedback probably promotes coherent and efficient transcription across the SBF/MBF regulon.

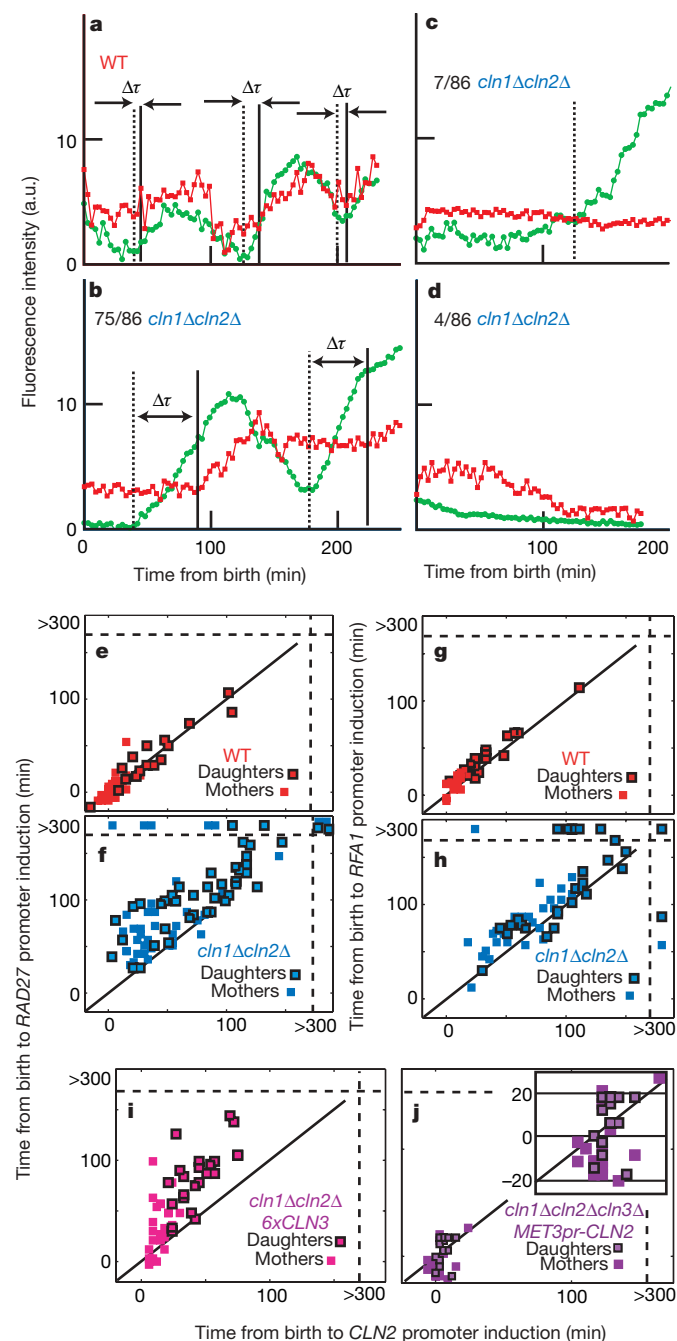
Further comparison of these three promoters in *cln1Δ cln2Δ* cells reveals that *CLN2* is almost always the first of the three to be activated, whereas the times to subsequent *RFA1pr* and *RAD27pr* inductions are significantly different from each other ( $P = 0.004$ ; Supplementary Table 3). This suggests that the *CLN2* promoter is the easiest for Cln3 to induce, followed by the *RFA1* promoter and then the *RAD27* promoter. We note that two MBF targets<sup>27–29</sup>, *RAD27* and *RFA1*, show different induction timing.

To address whether the lack of coherence in *cln1Δ cln2Δ* cells simply comes from low G1 cyclin levels, we analysed *cln1Δ cln2Δ* 6×*CLN3* cells (containing an extra five tandem integrated copies of *CLN3*). Although the expression of both the *CLN2* and *RAD27* promoters was significantly accelerated by extra *CLN3*, these cells still showed strongly incoherent expression compared to wild type (Fig. 2i).

To directly short-circuit the proposed positive feedback loop, we examined gene expression in *cln1Δ cln2Δ cln3Δ MET3pr-CLN2* cells on methionine-free medium (*MET3pr-CLN2* on). Although induction of *CLN2pr-GFP* and *RAD27-mCherry* was strongly accelerated by constitutive *CLN2* expression, incoherent expression compared to wild type was still observed (Fig. 2j). Notably, this incoherence was due to *RAD27-mCherry* induction before *CLN2pr-GFP*, compared to nearly simultaneous expression in wild type ( $-8 \pm 2$  min compared



to  $2 \pm 1$  min (mean  $\pm$  s.e.m.);  $P < 10^{-3}$ ), perhaps owing to differential loading of SBF (*CLN2*) and MBF (*RAD27*) regulated genes<sup>21,27–30</sup>.

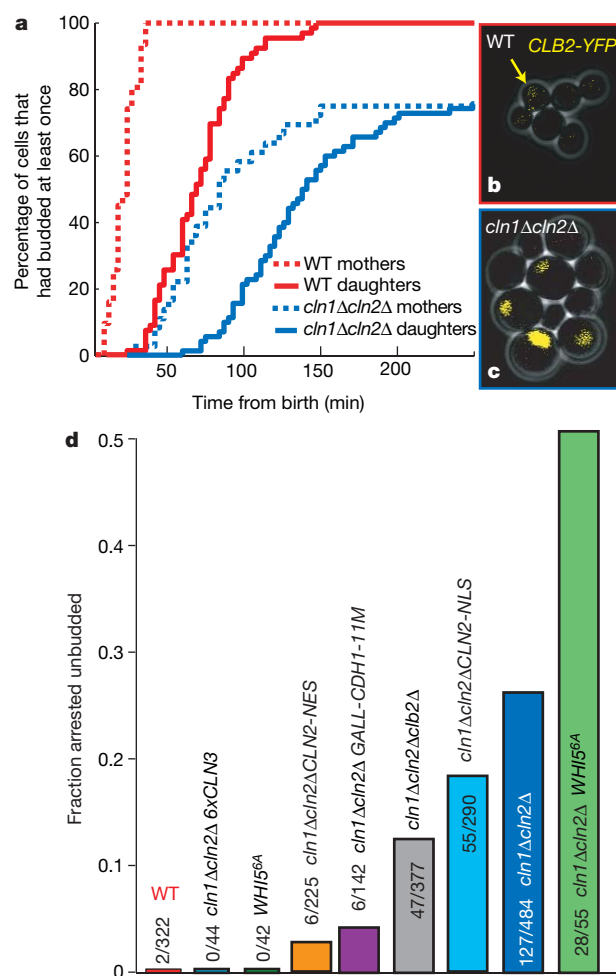


**Figure 2 | *Cln1* and *Cln2* drive coherent expression of the SBF/MBF regulon.** **a–f**, Strains containing both *CLN2pr-GFP* (green) and *RAD27-mCherry* (red) were examined (see Supplementary Information);  $\tau$  marks the computed time between *CLN2* and *RAD27* inductions. In wild type (**a**, **e**), all cells transcribed both markers synchronously; in *cln1Δ cln2Δ* (**b–d**, **f**), 75 cells transcribed both markers, with variable intervening intervals; 7 cells transcribed *CLN2pr-GFP* but not *RAD27-mCherry*; and 4 cells transcribed neither. Correlation of the initiation of *RAD27* and *CLN2* transcription in wild type (**e**) and *cln1Δ cln2Δ* (**f**) is shown; points beyond the dotted lines in (**f**, **h**) represent no transcription within 300 min (movie limit; see also Supplementary Table 3). **g**, **h**, Substituting *RFA1-mCherry* for *RAD27-mCherry* yielded similar results. **i**, **j**, *cln1Δ cln2Δ 6xCLN3* cells (**i**) and *cln1Δ cln2Δ cln3Δ* cells expressing *CLN2* from a *MET3* promoter (**j**) exhibited incoherent regulon expression compared to wild type, although expression of both *CLN2pr* and *RAD27pr* were faster than in *cln1Δ cln2Δ*.  $P < 10^{-3}$  for all comparisons.

Overall, these experiments suggest that the positive feedback architecture is a particularly effective way to promote coherent regulon expression.

### Stochastic cell cycle arrest

In addition to showing incoherent gene expression, 26% of *cln1Δ cln2Δ* cells fail to bud (Fig. 3a). We hypothesized that incoherent gene expression has a role in this sporadic unbudded arrest. Twenty out of 143 assayed *cln1Δ cln2Δ* cells were ‘strongly incoherent’: they failed to transcribe one or both of their two transcriptional markers (Fig. 2f, h); 90% of the strongly incoherent cells arrested unbudded, compared to 26% of all *cln1Δ cln2Δ* cells ( $P < 0.003$ ; Fig. 3a). Thus, a lack of coherence in the SBF/MBF regulon is a strong predictor of unbudded arrest within the *cln1Δ cln2Δ* population. *6xCLN3* reduced unbudded arrest in *cln1Δ cln2Δ* cells, perhaps because of accelerated regulon expression (Fig. 2i). Thus, unbudded arrest in *cln1Δ cln2Δ* cells may result from highly delayed expression of some regulon members.



**Figure 3 | Stochastic unbudded arrest in *cln1Δ cln2Δ* cells and its modulation by *Cln2*, *Cln3*, *Whi5*, and the mitotic cyclins.** **a**, A cumulative plot of the percentage of cells that budded at least once; 26% of *cln1Δ cln2Δ* cells arrest unbudded. **b**, **c**, Wild-type (**b**) and *cln1Δ cln2Δ* (**c**) cells with integrated *CLB2-YFP* fusion protein (endogenous promoter). Note that there is high nuclear *Clb2* specifically in large unbudded (arrested) *cln1Δ cln2Δ* cells. **d**, Delaying (*cln1Δ cln2Δ clb2Δ*) or removing (*cln1Δ cln2Δ GALL-CDH1-11M*) mitotic cyclin accumulation reduced the fraction of arrested cells; the addition of five copies of *CLN3* (*6xCLN3*) eliminated this arrest, whereas the addition of *WHI5<sup>6A</sup>* exacerbated the arrest. Unbudded arrest was weakly rescued by nuclear *Cln2* (*CLN2-NLS*), and strongly rescued by cytoplasmic *Cln2* (*CLN2-NES*). Unless stated otherwise in the text,  $P < 10^{-3}$  for all comparisons.

We hypothesized that in strongly incoherent cells, activation of only some regulon members might lead to activation of mitotic Clbs, which would then inactivate further SBF/MBF-regulated expression<sup>20</sup> (Fig. 1a and Supplementary Fig. 9). If the genes required for budding in the absence of *CLN1* and *CLN2*, such as *PCL1* and *PCL2* (ref. 31), had not yet been expressed by the time of Clb activation, unbudded arrest might ensue. Indeed, 20 out of 20 arrested *cln1Δ cln2Δ* cells contained large amounts of nuclear Clb2–yellow fluorescent protein (YFP) (Fig. 3b, c).

To test the role of transcription in unbudded arrest further, we deleted the rate-limiting SBF inhibitor *CLB2* in a *MET3pr-CLN2 cln1Δ cln2Δ* strain and observed a decrease in unbudded arrest from 26% to 13% (Fig. 3d). Additionally, we integrated unphosphorylatable Cdh1 under galactose control (*GALL-HA3-CDH1-m11*) into a *cln1Δ cln2Δ MET3pr-CLN2* strain to induce rapid degradation of all mitotic cyclins on galactose induction<sup>32</sup>. This reduced the unbudded arrested fraction to 4% in the first cell cycle after galactose induction (Fig. 3d). Because the timing of *CLB2pr-GFP* induction in *cln1Δ cln2Δ* cells was similar whether they arrested or not ( $P = 0.91$ ), the unbudded arrest was not due to unusually early *CLB2* induction.

Thus, mitotic cyclins promote unbudded arrest specifically in highly incoherent *cln1Δ cln2Δ* cells, probably owing to insufficient regulon expression before Clb-dependent SBF/MBF inactivation.

### Cln1 and Cln2 inactivate the transcriptional inhibitor WHI5

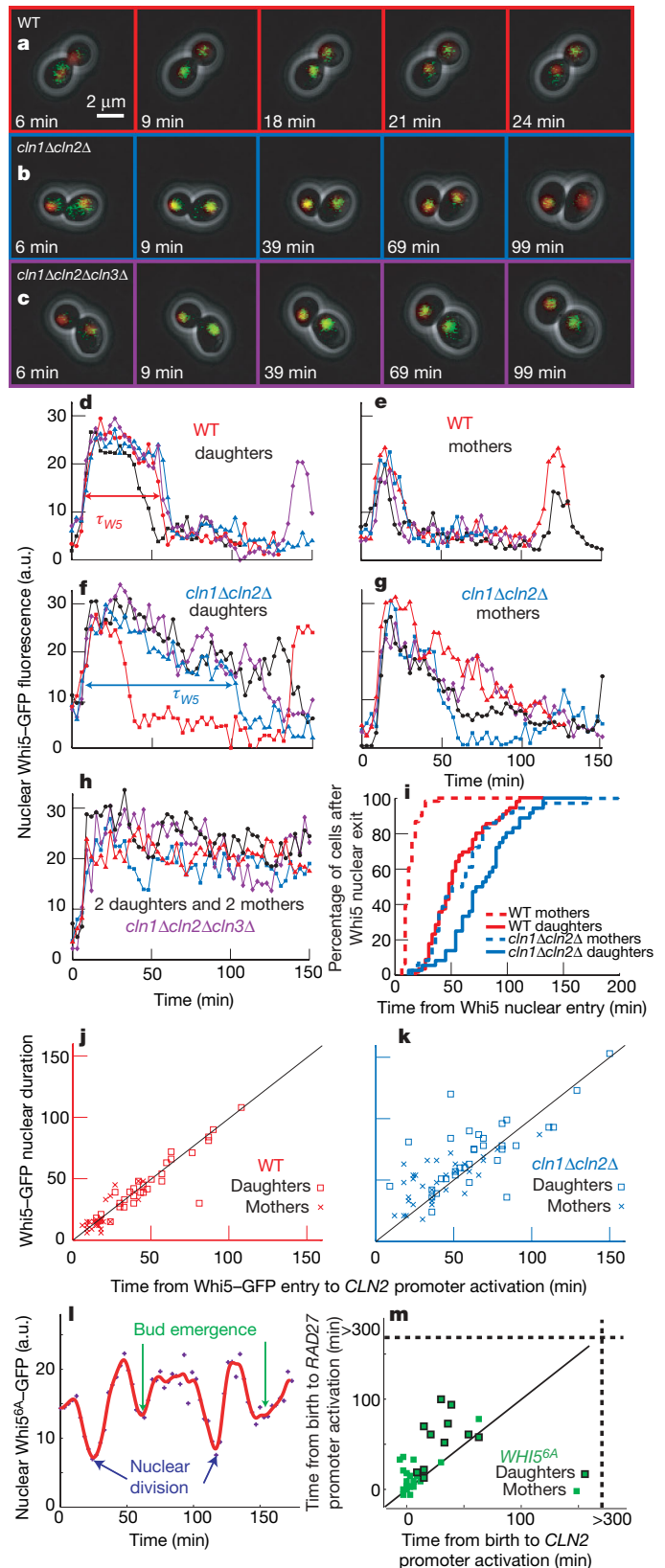
We wanted to determine if Cln1- and Cln2-dependent positive feedback operated through Whi5, a transcriptional inhibitor of the G1/S regulon<sup>18,19</sup>. Whi5 inactivation is rate-limiting for *CLN2* transcription and occurs by means of Cln-dependent phosphorylation, which leads to nuclear exclusion<sup>19</sup>.

First, we developed a quantitative assay for nuclear levels of Whi5–GFP by marking the nucleus with *HTB2-mCherry* (histone H2B) and measuring the difference between nuclear and cytoplasmic GFP fluorescence intensity (Fig. 4a–c). Whi5 entered the nucleus rapidly in both wild-type and *cln1Δ cln2Δ* cells. In wild-type cells, Whi5 also exited very rapidly; in *cln1Δ cln2Δ* cells, Whi5 exited much more slowly (Fig. 4d–g, i) consistent with biochemical data showing that Whi5 remains on the *CLN2* promoter longer in *cln1Δ cln2Δ* than in wild-type cells<sup>18</sup>. Because Whi5–GFP remained nuclear in *cln1Δ cln2Δ cln3Δ* cells (Fig. 4h), the slow Whi5 exit in *cln1Δ cln2Δ* cells is Cln3-dependent (this also excludes photobleaching artefacts). Thus, Cln3 initiates the slow exit of Whi5 from the nucleus, whereas Cln1 and Cln2 rapidly remove the remainder.

Because Whi5 exit and *CLN2* induction are tightly correlated in wild-type cells (Fig. 4j), translocation occurs shortly after Whi5 inactivation and coincides with activation of transcriptional positive feedback. *CLN2* promoter activation and Whi5 exit were less tightly correlated in *cln1Δ cln2Δ* cells consistent with the gradual exit of Whi5 (Fig. 4k and Supplementary Figs 5 and 6).

**Figure 4 | Cln1 and Cln2 are required for rapid phosphorylation and inactivation of the rate-limiting inhibitor Whi5.** **a–c**, Combined phase and fluorescence images showing Whi5–GFP and Htb2–mCherry (to mark the nucleus) fusion proteins for wild-type (**a**), *cln1Δ cln2Δ* (**b**) and *cln1Δ cln2Δ cln3Δ* (**c**) cells. The difference between nuclear and non-nuclear fluorescence intensity was used to quantify nuclear Whi5 by automated image analysis. **d–h**, Nuclear Whi5–GFP fluorescence. In comparison to wild-type cells (**d**, **e**), *cln1Δ cln2Δ* cells display delayed and less sharp Whi5 nuclear exit (**f**, **g**). Whi5 remains nuclear in *cln1Δ cln2Δ cln3Δ* cells (**h**). **i**, The percentage of cells in which Whi5 has left the nucleus (defined as attaining half the maximum amount) versus the time from Whi5 nuclear entry. **j**, **k**, Whi5 nuclear exit is tightly correlated with *CLN2* promoter activation in wild-type cells and less correlated in *cln1Δ cln2Δ* cells (see also Supplementary Table 3). **l**, Whi5<sup>6A</sup>–GFP (ref. 19), lacking 6 out of 12 Cln-dependent phosphorylation sites, reproducibly displayed significant, but slower and incomplete, shuttling out of the nucleus at Start and again at nuclear division. **m**, In *WHI5*<sup>6A</sup> strains containing *CLN2pr-GFP* and *RAD27-mCherry*, *CLN2* and *RAD27* induction were incoherent, correlating with the poor nuclear transport of Whi5<sup>6A</sup>–GFP.

To examine the role of Whi5 phosphorylation in positive feedback and regulon coherence, we used a *WHI5*<sup>6A</sup> allele<sup>19</sup> lacking 6 of the 12 Cln-dependent phosphorylation sites. Although Whi5<sup>6A</sup> was reported to be constitutively nuclear<sup>19</sup>, we observed significant, but slower and incomplete, shuttling of Whi5<sup>6A</sup>–GFP out of the nucleus at Start and again at nuclear division (10 out of 10 cells; Fig. 4l). *CLN2* and *RAD27* induction are less coherent in *WHI5*<sup>6A</sup> than in wild type (Fig. 4m; but





more coherent than *cln1Δ cln2Δ*), correlating with the poor nuclear transport of Whi5<sup>6A</sup>. Thus, interfering with the positive feedback loop by reducing the ability of Cln proteins to phosphorylate Whi5 reduces regulon coherence, even with all three G1 cyclins present.

The addition of WHI5<sup>6A</sup> to *cln1Δ cln2Δ* cells increased the frequency of unbudded arrest from 26% to 51%, consistent with the idea that unbudded arrest is a consequence of incoherent regulon expression in *cln1Δ cln2Δ* cells.

Overall, these results indicate that Whi5 is a Cln1 and Cln2 substrate in wild-type cells, and that this phosphorylation contributes to positive feedback. To determine whether Whi5 was the only such substrate, we compared timing of *CLN2* promoter activation for *whi5Δ* and *cln1Δ cln2Δ whi5Δ* cells (Supplementary Fig. 14 and Supplementary Table 3). Deletion of *WHI5* advances *CLN2* promoter induction in both wild-type and *cln1Δ cln2Δ* cells. Because *cln1Δ cln2Δ whi5Δ* cells delayed *CLN2pr* induction relative to *CLN1 CLN2 whi5Δ* cells, Cln1 and Cln2 probably act through Whi5-dependent and -independent mechanisms to promote positive feedback. Previous results indicated a Whi5-independent Cln3 requirement for SBF activation<sup>19</sup>, possibly acting through Swi6 (refs 19 and 33); a similar mechanism may be used by Cln1 and Cln2.

### Separable Cln2 functions

Cln1 and Cln2 are pleiotropic effectors of Start that have important nuclear and cytoplasmic functions<sup>34,35</sup>, complicating the interpretation of *cln1Δ cln2Δ* phenotypes. Therefore, we tested forced-localization *CLN2* alleles, expressed from the wild-type *CLN2* promoter, that restrict Cln2 to either the nucleus (*CLN2-NLS*) or the cytoplasm (*CLN2-NES*)<sup>34</sup>. *cln1Δ cln2Δ CLN2-NLS* cells show coherent regulon expression ( $P = 0.45$  compared to wild type), but *cln1Δ cln2Δ CLN2-NES* cells are highly incoherent compared to wild

type ( $P < 10^{-7}$ ); this indicates that coherent gene expression is primarily a nuclear function of *CLN2* (Fig. 5a, b (compare to Fig. 2), and Supplementary Table 3).

Consistent with a role of cytoplasmic Cln2 in budding<sup>34,35</sup>, integration of *CLN2-NES* into *cln1Δ cln2Δ* cells strongly reduces arrest (to 3%) in spite of less coherent gene expression. Furthermore, exogenous expression of *CLN2* drives cell cycle progression in previously blocked *cln1Δ cln2Δ* cells (Supplementary Fig. 10) and restores viability of *mbp1Δ swi4Δ* cells, which lack SBF and MBF and have very low regulon expression<sup>36,37</sup>. The localization mutants also have different efficacy for relieving unbudded arrest. Integration of *CLN2-NLS* into *cln1Δ cln2Δ* cells, providing coherent gene expression, led to a partial but significant reduction of unbudded arrest (from 26% to 19%;  $P = 0.04$ ).

Thus, cell morphogenesis and budding can be driven by two partially redundant pathways: by cytoplasmic Cln1 and Cln2 (refs 34 and 38), or by other genes in the G1/S regulon such as *PCL1* and *PCL2* (ref. 31; Fig. 5c). Having Cln1 and Cln2 coherently activate the G1/S regulon and directly drive bud emergence provides a compact solution to ensure efficient and timely morphogenesis and G1/S regulon expression, before subsequent Clb activation.

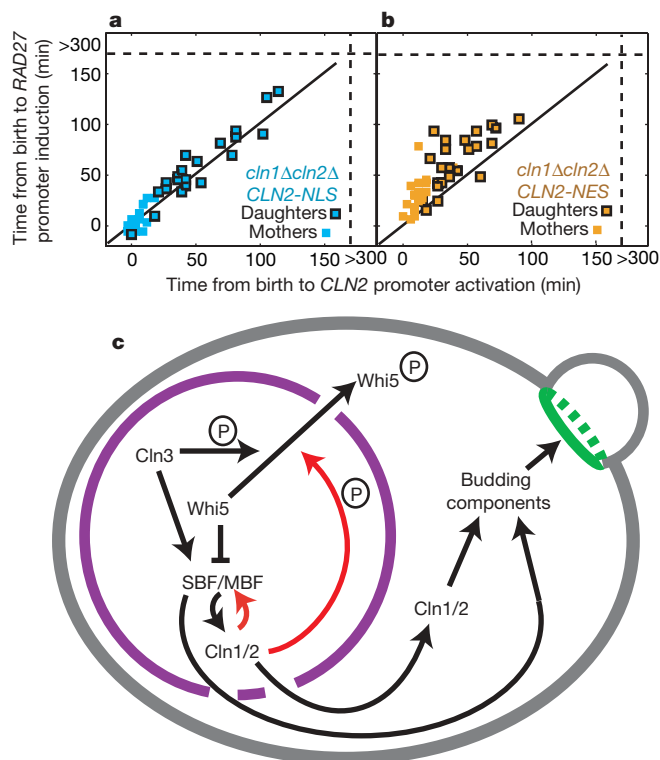
### Discussion

The regulatory architecture of the G1/S regulon provides an effective design to promote coordinated activation. The promoters are pre-loaded during G1 with a complex of factors that are subsequently rapidly activated by phosphorylation<sup>17,24,30</sup>, removing a potentially rate-limiting step. Furthermore, the upstream cyclin Cln3 is intrinsically more capable of triggering the *CLN2* promoter compared to two other randomly selected promoters from the regulon (*RFA1* or *RAD27*; Fig. 2e–h). The high sensitivity of *CLN1* and *CLN2* to Cln3 means that positive feedback from the initial burst of Cln1 and Cln2 will ensure that all other genes fire together. Indeed, in our experiments in wild-type cells, the genes are expressed too synchronously to evaluate which comes first. We find it probable that positive feedback will be a recurring motif in genetic control networks responsible for the coherent temporal coordination of multiple downstream events.

The sharpness of the Start switch, defined by the rapid exclusion of the transcriptional inhibitor Whi5 and the coherent expression of the G1/S regulon, is principally due to *CLN1*- and *CLN2*-dependent positive feedback (Fig. 5c, red lines) rather than a linear Cln3–Whi5–SBF pathway<sup>14,15,18,19</sup>. Our data are inconsistent with the sharpness of Start being primarily due to nonlinear increases in *CLN3* translation<sup>39</sup> or nuclear translocation<sup>40</sup>, or cooperative phosphorylation of Whi5 by Cln3 (ref. 19), because these mechanisms all predict a sharp switch in feedback-free *cln1Δ cln2Δ* cells.

In budding yeast, Start is a fundamental point of commitment at which physiological inputs such as nutrients, mating factor, size and cell type are integrated to produce an all-or-none decision. We show here that positive feedback provides robust switch-like cell cycle entry. Our single-cell data suggest that the point of commitment to the cell cycle, Start, is a very brief interval coinciding with the initiation of positive feedback and Whi5 exclusion. Subsequent Cln-dependent events, such as Sic1 phosphorylation and degradation<sup>41</sup> leading to DNA replication, could then be viewed as dependent on, rather than part of, Start.

This work also provides a molecular basis for understanding the modular structure of G1 (ref. 11). Two temporally uncorrelated processes in G1 are separated by the molecular event of Whi5 inactivation and nuclear exit. The upstream module is responsible for cell size control, whereas the downstream size-independent module actuates cell cycle progression<sup>11</sup>. Here, we showed that rapid Whi5 exit coincided with initiation of Cln1- and Cln2-dependent positive feedback. Once feedback is initiated, the rapidly accumulating Cln1 and Cln2 probably dominate cellular Cln-kinase activity, and Cln3, the rate-limiting upstream activator, is rendered unimportant. In general, we expect modularity, best shown by single-cell analysis, to be a signature of feedback-driven cellular control networks.



**Figure 5 | Function of nuclear Cln2 and a model for Start regulation by positive feedback.** **a**, **b**, Comparison of *cln1Δ cln2Δ* cells with either a nuclear localized (**a**) or a nuclear excluded (**b**) *CLN2* allele suggests that nuclear Cln2 is necessary and sufficient for regulon coherence. Strains contained *CLN2pr-GFP* and *RAD27-mCherry*. **c**, Model for regulon activation and bud emergence; red lines indicate pathways generating positive feedback.

Our systems-level analysis of Start provides a template for further studies of other checkpoints in yeasts or the G1/S transition in mammals. The utility of feedback at Start leads us to expect similar regulatory architecture across eukaryotes, even if the enabling genes are not homologous.

## METHODS SUMMARY

**Strain and plasmid constructions.** Standard methods were used throughout. All strains are W303-congenic.

**Time-lapse microscopy.** Preparation of cells for time-lapse microscopy was performed as previously described<sup>24</sup>. We integrated *MET3pr-CLN2* to conditionally express *Cln2* (ref. 14). By pre-growing cells without methionine before plating on media containing methionine (*MET3pr-CLN2* off), we were able to begin our time-lapse imaging experiments with similarly sized wild-type and *cln1Δ cln2Δ* cells. We imaged the first Start in cells that were budded at the time of transfer and that divided at least 30 min after methionine addition, to allow degradation of *Cln2* (refs 13 and 42) synthesized before *MET3* promoter turn-off.

**Image analysis.** Automated image segmentation and fluorescence quantification of yeast grown under time-lapse conditions were performed as previously described<sup>11,24</sup>. We added a function to previously described custom software<sup>24</sup> to identify nuclei labelled with Htb2–mCherry (histone H2B).

**Data analysis.** Fluorescence time series were extracted from movies as previously described<sup>24</sup>. Time series were fit using smoothing splines (Matlab) with a smoothing parameter of 0.001. We defined the onset of transcription for a G1/S fluorescent reporter by the maximum in the second derivative that fell between birth and budding (scored separately), which accurately locates rate changes in spite of noisy data and slow changes in the background fluorescence (Supplementary Figs 3 and 4).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 15 January; accepted 29 May 2008.

- Simchen, G., Pinon, R. & Salts, Y. Sporulation in *Saccharomyces cerevisiae*: premeiotic DNA synthesis, readiness and commitment. *Exp. Cell Res.* **75**, 207–218 (1972).
- Nachman, I., Regev, A. & Ramanathan, S. Dissecting timing variability in yeast meiosis. *Cell* **131**, 544–556 (2007).
- Shenhar, G. & Kassir, Y. A positive regulator of mitosis, Sok2, functions as a negative regulator of meiosis in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **21**, 1603–1612 (2001).
- Ferrell, J. E. Jr & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280**, 895–898 (1998).
- Xiong, W. & Ferrell, J. E. Jr. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426**, 460–465 (2003).
- Sha, W. *et al.* Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc. Natl Acad. Sci. USA* **100**, 975–980 (2003).
- Pomeroy, J. R., Sontag, E. D. & Ferrell, J. E. Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biol.* **5**, 346–351 (2003).
- Hartwell, L. H., Culotti, J., Pringle, J. R. & Reid, B. J. Genetic control of the cell division cycle in yeast. *Science* **183**, 46–51 (1974).
- Johnston, G. C., Pringle, J. R. & Hartwell, L. H. Coordination of growth with cell division in the yeast *Saccharomyces cerevisiae*. *Exp. Cell Res.* **105**, 79–98 (1977).
- Lord, P. G. & Wheals, A. E. Variability in individual cell cycles of *Saccharomyces cerevisiae*. *J. Cell Sci.* **50**, 361–376 (1981).
- Di Talia, S., Skotheim, J. M., Bean, J. M., Siggia, E. D. & Cross, F. R. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947–951 (2007).
- Jorgensen, P. & Tyers, M. How cells coordinate growth and division. *Curr. Biol.* **14**, R1014–R1027 (2004).
- Tyers, M., Tokiwa, G. & Futcher, B. Comparison of the *Saccharomyces cerevisiae* G<sub>1</sub> cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *EMBO J.* **12**, 1955–1968 (1993).
- Dirick, L., Bohm, T. & Nasmyth, K. Roles and regulation of Cln–Cdc28 kinases at the start of the cell cycle of *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4803–4813 (1995).
- Stuart, D. & Wittenberg, C. CLN3, not positive feedback, determines the timing of CLN2 transcription in cycling cells. *Genes Dev.* **9**, 2780–2794 (1995).
- Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
- Kato, M., Hata, N., Banerjee, N., Futcher, B. & Zhang, M. Q. Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol.* **5**, R56 (2004).
- de Bruin, R. A., McDonald, W. H., Kalashnikova, T. I., Yates, J. III & Wittenberg, C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell* **117**, 887–898 (2004).
- Costanzo, M. *et al.* CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell* **117**, 899–913 (2004).
- Amon, A., Tyers, M., Futcher, B. & Nasmyth, K. Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell* **74**, 993–1007 (1993).
- de Bruin, R. A. *et al.* Constraining G1-specific transcription to late G1 phase: the MBF-associated corepressor Nrm1 acts via negative feedback. *Mol. Cell* **23**, 483–496 (2006).
- Cross, F. R. & Tinkelenberg, A. H. A potential positive feedback loop controlling CLN1 and CLN2 gene expression at the start of the yeast cell cycle. *Cell* **65**, 875–883 (1991).
- Dirick, L. & Nasmyth, K. Positive feedback in the activation of G1 cyclins in yeast. *Nature* **351**, 754–757 (1991).
- Bean, J. M., Siggia, E. D. & Cross, F. R. Coherence and timing of cell cycle Start examined at single-cell resolution. *Mol. Cell* **21**, 3–14 (2006).
- Mateus, C. & Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313–1323 (2000).
- Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. *Sci. STKE* **2006**, re17 (2006).
- Iyer, V. R. *et al.* Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
- Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
- Simon, I. *et al.* Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
- Koch, C., Schleiffer, A., Ammerer, G. & Nasmyth, K. Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at Start, whereas Clb/Cdc28 kinases displace it from the promoter in G<sub>2</sub>. *Genes Dev.* **10**, 129–141 (1996).
- Moffat, J. & Andrews, B. Late-G1 cyclin-CDK activity is essential for control of cell morphogenesis in budding yeast. *Nature Cell Biol.* **6**, 59–66 (2004).
- Zachariae, W., Schwab, M., Nasmyth, K. & Seufert, W. Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science* **282**, 1721–1724 (1998).
- Wijnen, H., Landman, A. & Futcher, B. The G<sub>1</sub> cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6. *Mol. Cell Biol.* **22**, 4402–4418 (2002).
- Edgington, N. P. & Futcher, B. Relationship between the function and the location of G1 cyclins in *S. cerevisiae*. *J. Cell Sci.* **114**, 4599–4611 (2001).
- Miller, M. E. & Cross, F. R. Distinct subcellular localization patterns contribute to functional specificity of the Cln2 and Cln3 cyclins of *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **20**, 542–555 (2000).
- Koch, C., Moll, T., Neuberger, M., Ahorn, H. & Nasmyth, K. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* **261**, 1551–1557 (1993).
- Bean, J. M., Siggia, E. D. & Cross, F. R. High functional overlap between MBF and SBF in the G1/S transcriptional program in *Saccharomyces cerevisiae*. *Genetics* **171**, 49–61 (2005).
- McCusker, D. *et al.* Cdk1 coordinates cell-surface growth with the cell cycle. *Nature Cell Biol.* **9**, 506–515 (2007).
- Polymenis, M. & Schmidt, E. V. Coupling of cell division to cell growth by translational control of the G<sub>1</sub> cyclin CLN3 in yeast. *Genes Dev.* **11**, 2522–2531 (1997).
- Wang, H., Gari, E., Verges, E., Gallego, C. & Aldea, M. Recruitment of Cdc28 by Whi3 restricts nuclear accumulation of the G1 cyclin–Cdk complex to late G1. *EMBO J.* **23**, 180–190 (2004).
- Schneider, B. L., Yang, Q. H. & Futcher, A. B. Linkage of replication to Start by the Cdk inhibitor Sic1. *Science* **272**, 560–562 (1996).
- Lanker, S., Valdivieso, M. H. & Wittenberg, C. Rapid degradation of the G<sub>1</sub> cyclin Cln2 induced by CDK-dependent phosphorylation. *Science* **271**, 1597–1601 (1996).
- Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22**, 1567–1572 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the National Institute of Health (J.M.S., E.D.S. and F.R.C.), the Burroughs Wellcome Fund (J.M.S.) and the National Science Foundation (E.D.S.). We thank N. Buchler, G. Charvin, B. Drapkin and J. E. Ferrell for conversations; J. Widom and C. Wittenberg for comments on the manuscript; J. M. Bean, B. Timney and J. Robbins for help with strain/plasmid construction; M. Schwab for the plasmid pWS358; B. Futcher for the *CLN2-NES* and *CLN2-NLS* plasmids; E. Bi for the pKT355 mCherry tagging plasmid; and M. Tyers for *WHI5* phosphorylation site mutant strains and plasmids.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.M.S. ([skotheim@stanford.edu](mailto:skotheim@stanford.edu)).



## METHODS

**Strain and plasmid constructions.** Standard methods were used throughout. All strains are W303-congenic. In synchronized wild-type cells, *GFP* mRNA from the *CLN2* promoter and *CLN2* mRNA follow similar kinetics, and accumulation of cellular fluorescence follows with a slight delay<sup>24</sup>. *WHI5*<sup>6A</sup> and *WHI5*<sup>6A</sup>-*GFP* strains with modified *WHI5* at the endogenous locus were a gift from M. Tyers. Plasmids for introduction of *CLN2*-*NES* and *CLN2*-*NLS* under control of the *CLN2* promoter were obtained from B. Futcher, and integrated at the *ura3* locus in a *cln1Δ cln2Δ* background. Histone H2B (*HTB2*) was carboxy-terminally tagged with *mCherry* using PCR-mediated tagging with the template plasmid pKT355 (ref. 43). *RAD27* and *RFA1* were tagged similarly. All other alleles were from laboratory stocks described previously.

**Time-lapse microscopy.** Preparation of cells for time-lapse microscopy was performed as previously described<sup>24</sup>. Because mutant cells are larger than wild type, we integrated *MET3pr-CLN2* to conditionally express Cln2 (ref. 14). On media lacking methionine (*MET3pr-CLN2* on), cells bud and divide at comparable sizes (Supplementary Fig. 3). By pre-growing cells without methionine before plating on media containing methionine (*MET3pr-CLN2* off), we were able to begin our time-lapse imaging experiments with similarly sized wild-type and *cln1Δ cln2Δ* cells. We imaged the first Start in cells that were budded at the time of transfer and that divided at least 30 min after methionine addition, to allow degradation of Cln2 (refs 13 and 42) that was synthesized before *MET3* promoter turn-off. In brief, growth of microcolonies was observed with fluorescence time-lapse microscopy at 30 °C using a Leica DMIRE2 inverted microscope with a Ludl motorized XY stage. Images were acquired every 3 min for cells grown in glucose and every 6 min for cells grown in glycerol/ethanol with a Hamamatsu Orca-ER camera. Custom Visual Basic software integrated with ImagePro Plus was used to automate image acquisition and microscope control.

**Image analysis.** Automated image segmentation and fluorescence quantification of yeast grown under time-lapse conditions were performed as previously described<sup>24</sup>. Budding was scored visually, and cell birth was scored by the disappearance of Myo1-GFP at the bud neck, generally with single-frame accuracy. Background was measured as the average fluorescence of unlabelled cells and subtracted from the measured pixel intensities. We added a function to previously described custom software<sup>24</sup> to identify nuclei labelled with Htb2-mCherry (histone H2B). The red signal was smoothed, disconnected fragments were eliminated and the cells with nuclei that were too small, dim or oddly shaped (area versus minimally enclosed rectangle) were eliminated. After background subtraction, the nucleus was defined to be where the fluorescence was greater than 70% of maximum, which controls for cell variability and vertical movement of the nucleus. The nuclear Whi5-GFP signal was the difference between the average nuclear and cytosolic intensities.

**Data analysis.** *P* values using appropriate tests yielded  $P < 0.001$  for all comparisons in the text, except where noted. Fluorescence time series were extracted from movies as previously described<sup>24</sup>. Time series were fit using smoothing splines (Matlab) with a smoothing parameter of 0.001. We defined the onset of transcription for a G1/S fluorescent reporter by the maximum in the second derivative that fell between birth and budding (scored separately). This method was chosen because it accurately locates rate changes in spite of noisy data and slow changes in the background fluorescence. The onset time was nearly unchanged over a range of  $10^3$  in smoothing parameter (Supplementary Figs 3 and 4).

# The cohesin ring concatenates sister DNA molecules

Christian H. Haering<sup>1\*†</sup>, Ana-Maria Farcas<sup>1\*</sup>, Prakash Arumugam<sup>1†</sup>, Jean Metson<sup>1</sup> & Kim Nasmyth<sup>1</sup>

**Sister chromatid cohesion, which is essential for mitosis, is mediated by a multi-subunit protein complex called cohesin. Cohesin's Scc1, Smc1 and Smc3 subunits form a tripartite ring structure, and it has been proposed that cohesin holds sister DNA molecules together by trapping them inside its ring. To test this, we used site-specific crosslinking to create chemical connections at the three interfaces between the three constituent polypeptides of the ring, thereby creating covalently closed cohesin rings. As predicted by the ring entrapment model, this procedure produced dimeric DNA–cohesin structures that are resistant to protein denaturation. We conclude that cohesin rings concatenate individual sister minichromosome DNA molecules.**

Sister chromatid cohesion is mediated by a multi-subunit complex called cohesin which contains four core subunits: Smc1 and Smc3, which are members of the structural maintenance of chromosomes (SMC) protein family, and two non-SMC subunits, Scc1 (also called Mcd1) which is a member of the kleisin family, and Scc3 (SA)<sup>1,2</sup>. Sister chromatid disjunction occurs when all chromosomes have been bi-oriented, and it is triggered by site-specific cleavage of the Scc1 subunit of cohesin by separase<sup>3</sup>. The Smc1 and Smc3 subunits of cohesin both form rod-shaped molecules that heterodimerize by means of 'hinge' domains situated at the ends of 30-nm-long intramolecular antiparallel coiled coils<sup>4,5</sup>. ATPase 'heads' at the other ends are connected by the Scc1 kleisin subunit of cohesin, thereby forming a tripartite ring with a 35 nm diameter<sup>4,6</sup>. It has been proposed that cohesin holds sister chromatids together by trapping sister DNAs inside its ring. By severing Scc1, separase is thought to open the ring and thereby release sister DNAs from their topological embrace.

To investigate the physical nature of sister chromatid cohesion, we recently used sucrose gradient sedimentation and gel electrophoresis to purify cohesed (held together by cohesin) sister chromatids of small circular minichromosomes from yeast<sup>7</sup>. The minichromosome dimers are composed of individual DNAs packaged into nucleosomes that are converted to monomers by cleaving Scc1 or by linearizing their DNA. Notably, their formation depends on centromeres as well as cohesin (see Supplementary Information and Supplementary Fig. 1). These data are consistent with (but do not prove) the notion that cohesin attaches to chromatin using a topological mechanism. However, they do not exclude the possibility that cohesion requires the non-topological association of cohesin rings bound to different chromatin fibres<sup>8</sup>. If cohesin holds dimeric minichromosomes together by trapping them inside its ring, then introducing covalent connections between the Smc1–Smc3 hinge, Smc1 head–Scc1-carboxy terminal and Smc3 head–Scc1-amino terminal interfaces should create a chemically circularized cohesin ring within which sister DNAs would be trapped even after protein denaturation (Fig. 1a). We describe here experiments that test this prediction.

## Covalent connection of the three cohesin ring subunits

To connect cohesin subunit interfaces covalently, we used the homobifunctional thiol-reactive chemicals dibromobimane (bBBR) and

bis-maleimidoethane (BMOE) which bridge thiol groups up to 5 Å or 8 Å respectively<sup>9,10</sup> (Supplementary Fig. 2). We created a homology model of the yeast Smc1–Smc3 hinge heterodimer based on the homodimeric *Thermotoga maritima* crystal structure<sup>4,11</sup> and identified two juxtaposed side chains that we mutated to cysteines. Incubation of the engineered Smc1–Smc3 hinge dimer with either bBBR or BMOE caused efficient crosslinking within a few minutes (Fig. 1b). Notably, crosslinking was dependent on both cysteine substitutions. We used the same approach to connect the loop between the two  $\beta$ -strands of the winged-helix of Scc1 to a  $\beta$ -strand in the ATPase head of Smc1 (ref. 12; Fig. 1c). Because no structural information is available for the interface between the ATPase head of Smc3 and the amino-terminal domain of Scc1, we expressed Smc3 and Scc1 as a fusion protein, using a long flexible linker containing triple target sequences for the TEV protease to connect the C terminus of Smc3 with the N terminus of Scc1 (Smc3–TEV–Scc1)<sup>13</sup>. To create cohesin rings that could be chemically circularized by BMOE or bBBR, the cysteine substitutions were introduced into both Smc3–TEV–Scc1 and Smc1.

## Crosslinking produces SDS-resistant minichromosome dimers

A 2.3-kilobase (kb) circular minichromosome<sup>7</sup> was introduced into yeast strains in which the Smc1 and Smc3–TEV–Scc1 polypeptides contained either all four cysteine substitutions or only a subset of these. After nocodazole arrest and cell lysis, extracts were centrifuged through sucrose gradients, and fractions containing monomeric and dimeric minichromosomes were detected by native agarose gel electrophoresis and Southern blotting. Dimeric minichromosomes could still be isolated from yeast cells in which the cohesin ring subunits had been engineered to permit cohesin circularization. The cysteine substitutions had little adverse effect, but the fusion of Scc1 to Smc3 roughly halved the fraction of dimeric minichromosomes (Fig. 2a). This was not surprising as the fusion causes partial cohesion defects *in vivo*<sup>13</sup>. Dithiothreitol (DTT), sucrose and other low-molecular-mass contaminants were removed from the gradient fractions by dialysis and cohesin subunits were treated with bBBR, BMOE or merely dimethylsulphoxide (DMSO) solvent. After quenching the reaction by the re-addition of DTT, SDS was added to a final concentration of 1% and the samples were heated to

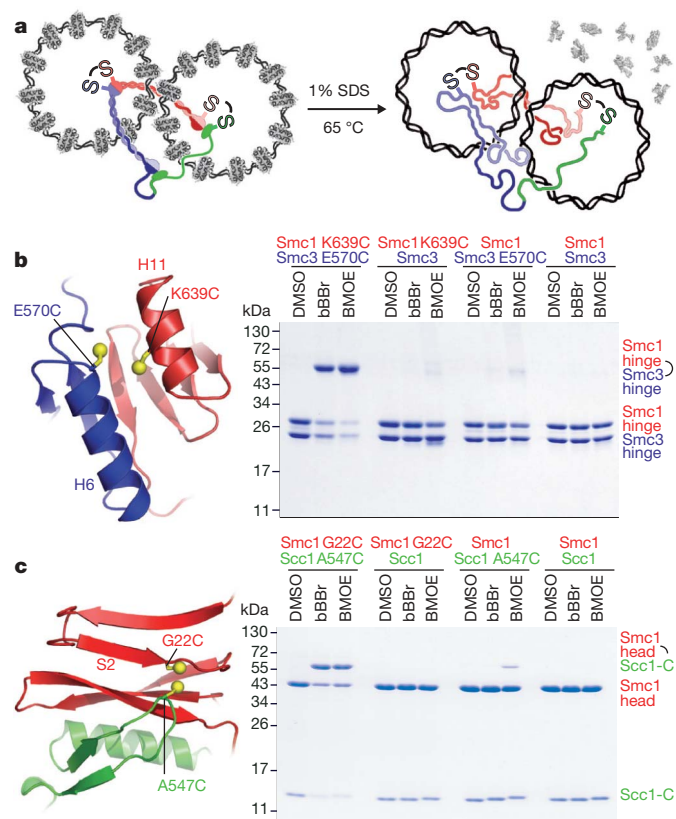
<sup>1</sup>University of Oxford, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK. <sup>†</sup>Present addresses: European Molecular Biology Laboratory (EMBL), Meyerhofstraße 1, 69117 Heidelberg, Germany (C.H.H.); Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK (P.A.).

\*These authors contributed equally to this work.

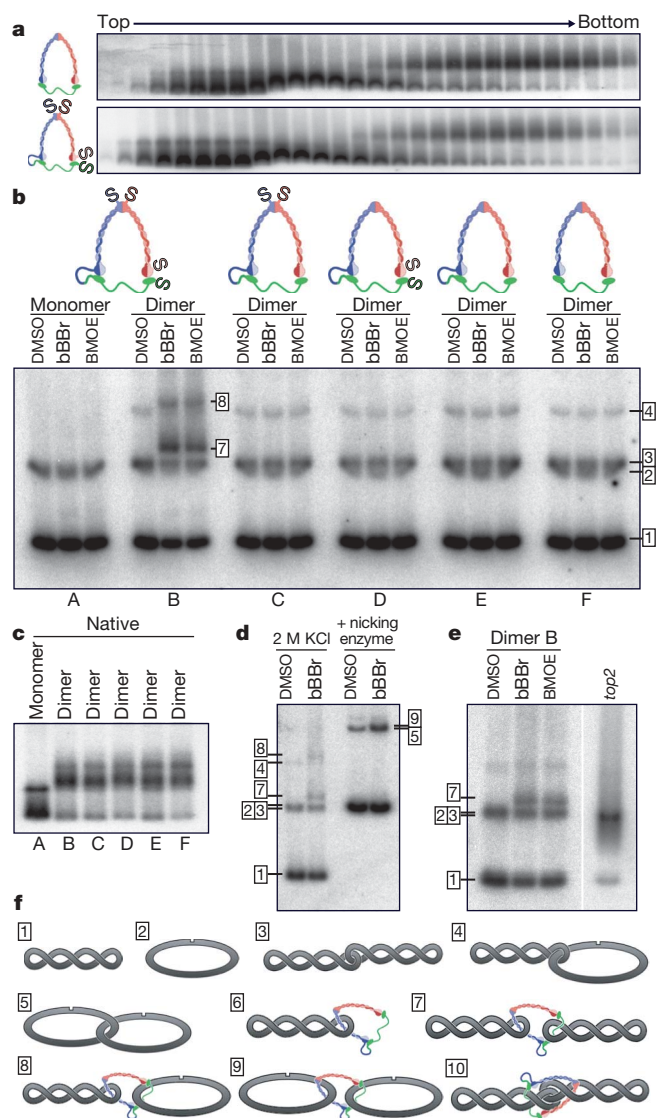


65 °C for 4 min. The denatured samples were finally electrophoresed in agarose gels containing ethidium bromide, and minichromosome DNA was detected by Southern blotting.

Dimer fractions from control cells expressing unmodified cohesin contained four species of DNA (Fig. 2b, panel F). The fastest migrating and most predominant are supercoiled monomers (species 1; Fig. 2f). Owing to SDS in the loading buffer, the next two species (2 and 3) were poorly resolved from each other. These DNAs co-migrated with monomeric nicked circles produced by nicking enzyme after removal of nucleosomes with 2 M potassium chloride (Fig. 2d) and with (infrequently) intertwined—that is, concatenated—supercoiled DNAs isolated from a topoisomerase II mutant (Fig. 2e), and therefore include both of these species of DNA. The nicking enzyme treatment revealed that about 10% of DNAs from dimer (but not monomer) fractions are DNA–DNA concatemers (Fig. 2d and data not shown). The least abundant species (4) migrated more slowly than two intertwined supercoiled circles (3) but more rapidly than two intertwined nicked circles generated by treatment with nicking enzyme (5). We conclude that these DNAs



**Figure 1 | Making covalently closed cohesin rings.** **a**, Fusion of the C terminus of Scc1 with the N terminus of Smc3 and chemical crosslinking of engineered cysteine residues at the Smc1–Smc3 and Smc1–Scc1 interfaces creates a covalently closed cohesin ring. **b**, A homology model of the Smc1–Smc3 hinge interface using the homodimeric bacterial structure (Protein Data Bank (PDB) 1GXL) identifies two juxtaposed C $\beta$  atoms in helix H11 of Smc1 and helix H6 of Smc3 at a distance compatible with crosslinking when mutated to cysteine. A Coomassie stained SDS–PAGE of wild-type and cysteine mutant yeast Smc1–Smc3 hinge domain dimers is shown. **c**, The structure of the yeast Scc1 C terminus bound to the head domain of Smc1 (PDB 1W1W) identifies two juxtaposed side chains that should allow crosslinking when mutated to cysteine. Complexes of wild-type and cysteine mutant yeast Smc1 head domain bound to the C-terminal domain of Scc1 resolved by SDS–PAGE are shown. The low-level crosslinking observed for the Smc1–Scc1(A547C) combination probably results from a reaction of BMOE with the engineered cysteine in Scc1 and the nearby  $\epsilon$ -amino group of Smc1's lysine residue K20.



**Figure 2 | Covalent cohesin circularization creates SDS-resistant minichromosome DNA dimers.** **a**, Extracts from yeast strains harbouring the minichromosome and expressing wild-type cohesin (top) or Smc3–TEV–Scc1 and Smc1 containing the engineered cysteine pairs (bottom) were separated by gradient centrifugation. Minichromosome dimers sediment faster but electrophorese slower than monomers. **b**, Minichromosome monomer or dimer gradient fractions from yeast strains in which all three (K14856; panels A and B), two (K14857, K14859; panels C and D), one (K14858; panel E) or no (K14860; panel F) ring subunit interface(s) can be covalently linked were treated with DMSO, bBBr or BMOE and separated after denaturation. Monomer and dimer fractions contain supercoiled (1) and nicked (2) monomeric minichromosomes. Dimer fractions also contain supercoiled–supercoiled (3) and supercoiled–nicked (4) concatenated minichromosome DNAs. Only samples in which all three cohesin ring subunit interfaces have been covalently linked contain additional slower migrating bands, presumably corresponding to supercoiled–supercoiled (7) and supercoiled–nicked (8) cohesed minichromosomes. **c**, Input fractions for crosslinking reactions A–F run without denaturation. **d**, Treatment of non-crosslinked or crosslinked minichromosomes with nicking enzyme, after removal of nucleosomes by high salt, converts supercoiled (1), supercoiled–supercoiled (3, 7) or supercoiled–nicked (4, 8) minichromosomes to nicked (2) or nicked–nicked (5, 9) forms. **e**, A gradient dimer fraction from a topoisomerase II mutant strain (K15029) grown at the restrictive temperature was denatured. Concatenated minichromosomes (short exposure) co-migrate with bands (2) and (3) of K14856 (long exposure) on the same gel. **f**, Schemata of minichromosome conformations.

correspond to one supercoiled circle intertwined with one nicked circle.

Treatment of dimer fractions with bBBR or BMOE had no effect on the electrophoresis profile of minichromosome DNAs (Fig. 2b, panel F). Moreover, the very same pattern was observed when dimeric minichromosomes were isolated from strains expressing the Smc3–TEV–Sccl fusion (Fig. 2b, panel E) and crosslinkable cysteine pairs at either the Smc1–Sccl or the Smc1–Smc3 interface (Fig. 2b, panels C and D). In contrast, bBBR and BMOE but not DMSO alone caused the appearance of two additional species of DNA when dimers were isolated from a strain containing the Smc3–TEV–Sccl fusion and cysteine pairs at both interfaces (Fig. 2b, panel B). The more abundant DNA species (7) migrated slightly more slowly than intertwined supercoiled circles, whereas the less abundant (8) migrated slightly more slowly than supercoiled circles intertwined with nicked circles. Their electrophoretic mobilities and the fact that neither was detected when identical crosslinking reactions were conducted with monomer fractions (Fig. 2b, panel A and Fig. 2c) suggest that they represent novel dimeric forms. Their formation occurs at the expense of supercoiled and nicked monomeric circles.

Our data suggest that the faster form (7) is a dimer of supercoiled monomeric circles associated with cohesin whereas the slower form (8) is a dimer between supercoiled and nicked circles associated with cohesin. Consistent with this, both were converted to a form (9) that co-migrates with intertwined nicked circles (5) when treated with nicking enzyme (Fig. 2d). The extra mass of cohesin probably has little effect on the electrophoretic mobility of this slow running species. Notably, neither novel dimer was produced when just one of the four cysteine substitutions was lacking (Supplementary Fig. 3a), suggesting that they arise due to the simultaneous crosslinking of both cysteine pairs at the Smc1–Smc3 hinge and Smc1–Sccl interfaces. Crosslinked dimers were also produced when minichromosomes

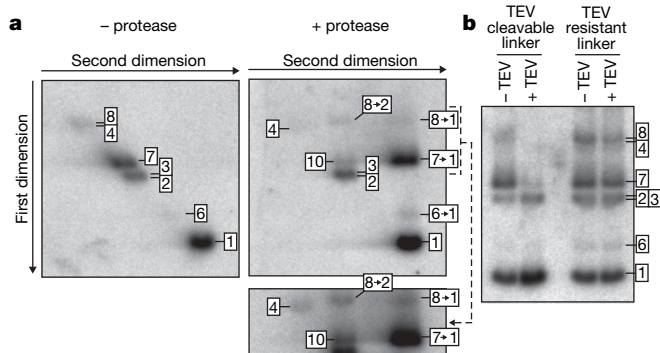
were isolated from cycling cultures (Supplementary Fig. 3a), suggesting that their formation is not an artefact caused by arresting cells with nocodazole. Finally, no slower migrating species could be observed when DNA was linearized with a restriction enzyme after crosslinking (Supplementary Fig. 3b). In conclusion, covalent closure of the cohesin ring converts dimeric but not monomeric minichromosomes to a dimeric form that is resistant both to SDS and to 2 M potassium chloride (native dimers are converted to monomers at 0.5–1 M potassium chloride; Supplementary Fig. 4).

### Circularized cohesin holds individual DNAs together

To test whether the SDS-resistant dimers produced by cohesin circularization are indeed monomeric DNAs held together by cohesin, we used two-dimensional gel electrophoresis. Denatured crosslinked samples were resolved on an agarose gel as before (the first dimension), and then electrophoresed perpendicularly through a thin zone of agarose or agarose containing proteinase K into a second agarose gel (the second dimension). Proteinase K should digest any proteins before DNAs enter the second gel and DNAs that ran as dimers in the first dimension should run as monomers in the second dimension if they were initially held together by a proteinaceous (that is, cohesin) connection. In the absence of proteinase K, all DNA species migrate identically in first and second dimensions and therefore lie on a diagonal line (Fig. 3a). Several species also ran on the diagonal in the presence of proteinase K, namely monomeric supercoils (1), monomeric nicked circles (2), intertwined supercoils (3), and nicked circles intertwined with supercoils (4) (Fig. 3a). In contrast, DNAs of presumptive dimers of two supercoiled minichromosomes held together by cohesin (7) migrated as monomeric supercoils in the second dimension (7→1), whereas presumptive supercoiled–nicked circle dimers held together by cohesin (8) split into monomeric supercoils (8→1) and nicked circles (8→2). Thus, chemical circularization of cohesin associated with native dimeric minichromosomes is accompanied by the crosslinking of monomeric DNAs to create SDS-resistant but protease-sensitive dimers.

The protease-containing two-dimensional gel revealed two new types of low abundance DNAs. The first (6) migrated considerably slower than monomeric supercoils in the first dimension but ran as monomeric supercoils in the second dimension (6→1). These DNAs were only detected in monomeric or dimeric minichromosome preparations in which cohesin rings had been covalently closed (data not shown). They presumably correspond to rare supercoiled monomers the migration of which is retarded by their association with (entrapment by) a chemically circularized cohesin ring. The second species (10) co-migrated with cohesin-mediated supercoiled dimers in the first dimension but with intertwined supercoils (and nicked circles) in the second dimension. These DNAs could correspond either to monomeric nicked circles associated with cohesin or, more probably, to intertwined supercoils that are also associated with cohesin.

If cohesin circularization by bBBR and BMOE crosslinking *per se* is responsible for the formation of SDS-resistant minichromosome dimers, then cleavage of the cohesin ring should be sufficient to release the monomeric DNAs. To test this, we incubated crosslinked dimeric minichromosome preparations with or without TEV protease to cleave the linker connecting Smc3 and Sccl. The presence of TEV greatly reduced both types of DNA dimers induced by the circularization of cohesin (7 and 8), which was accompanied by a corresponding increase in monomeric DNAs (1 and 2) (Fig. 3b). This effect was clearly caused by cleavage of the TEV sites in the Smc3–Sccl linker because DNA dimers produced by circularization of cohesin with a TEV-resistant Smc3–Sccl linker were unaffected by TEV protease (Fig. 3b). We conclude that the SDS-resistant association of sister DNAs induced by crosslinking the three subunits of cohesin does not merely accompany the circularization of cohesin but actually depends on it.



**Figure 3 | Covalent circularization of cohesin holds individual DNAs together.** **a**, Crosslinked and denatured samples were run on an agarose gel, excised and run on a second agarose gel in perpendicular direction. A thin slot between the first dimension lane and the second gel was filled with either agarose or agarose containing 0.2 mg ml<sup>-1</sup> proteinase K. In the absence of protease, all bands run on a diagonal. In the presence of protease, supercoiled–supercoiled cohesed dimers (7) run as supercoiled monomers (7→1) in the second dimension, whereas supercoiled–nicked cohesed dimers (8) split up into supercoiled (8→1) and nicked (8→2) monomers in the second dimension, more clearly visible in a longer exposure (bottom panel). A faint band, presumably corresponding to cohesin bound supercoiled monomers (6), is converted into naked supercoiled monomers (6→1) on protease cleavage. An additional band (10) running off the diagonal presumably corresponds to supercoiled–supercoiled concatenated minichromosomes that had bound cohesin. **b**, Crosslinked minichromosome samples were treated with or without TEV protease before denaturation. The slower migrating bands (7 and 8) corresponding to cohesed minichromosomes disappear on TEV incubation if the Smc3–Sccl linker contains triple target sequences for the protease (TEV-cleavable), but not if the –1 sites of the triple target sequence are mutated to lysine (TEV-resistant).



### Cohesion by single or double cohesin rings?

The simplest explanation for the crosslinking results is that sister DNAs are topologically trapped within single (monomeric) cohesin rings (Supplementary Fig. 5a). An alternative albeit more complicated possibility envisions entrapment of sister DNAs either by rings that are themselves topologically intertwined (Supplementary Fig. 5b) or by dimeric cohesin rings. Only two cysteine crosslinks are needed for entrapment by single rings, whereas four are required by double ring models. If we knew the efficiency with which bBBR and BMOE crosslink the Smc1–Smc3 and Smc1–Scc1 interfaces and the number of cohesin bridges, then we could calculate the fraction of DNAs that should be trapped as dimers according to the two models and compare these to what is actually observed.

To estimate the protein crosslinking efficiency, we added purified Smc1–Smc3 hinge or Smc1 head–Scc1–C to crosslinkable minichromosome dimer preparations before crosslinking with bBBR or BMOE and denaturation. One-half of the reaction was run on SDS–PAGE and the fraction of crosslinked proteins was measured after SYPRO Ruby staining (Fig. 4a). The other half was run on an agarose gel and the fraction of dimerized DNA was measured by Southern blotting (Fig. 4b). The fraction of rings expected to be crosslinked at both interfaces, which is given by multiplying individual crosslinking efficiencies, was 30% for both bBBR and BMOE. We would therefore expect 30% of DNAs to be dimerized if held together by a single cohesin ring but only 9% by a double ring. Estimating the actual number of bridges is harder because the gradient fractions contain much cohesin that is not associated with minichromosomes (data not shown). However, if we assume that a single bridge is sufficient to hold sister DNAs together and that cross-bridges form *in vivo* and survive fractionation *in vitro* with a defined probability ( $\lambda$ ), then the fraction of chromosomes  $f(x)$  with  $x$  bridges should fit a Poisson distribution.  $f(0)$  can be measured directly, namely by measuring the fraction of monomeric minichromosomes (Supplementary Fig. 6a) and DNA–DNA concatemers (Fig. 2d), which permits calculation of  $\lambda$  (see Supplementary Information). Because  $f(0)$  is large, most native dimeric minichromosomes are predicted to have a single bridge. Taking this into account, the single and double ring models predict 32% and 10% dimerization, respectively. The observed value with both reagents was 30% (Fig. 4b), which is inconsistent with the double ring model and close to that predicted by the single ring model.

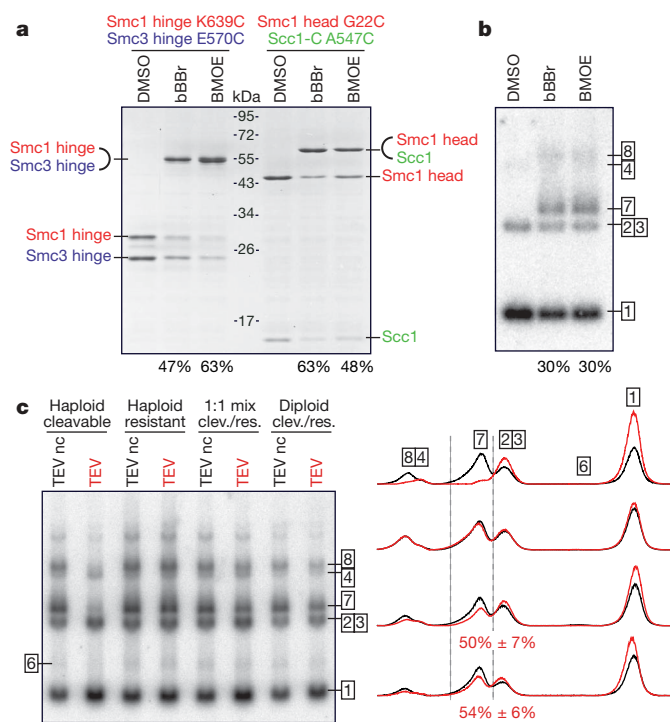
Single and double ring models also make different predictions for heterozygous diploids that express equal amounts of TEV-cleavable and TEV-resistant Smc3–Scc1 fusion proteins (Supplementary Fig. 6b). In the case of one cross-bridge, the single ring model predicts that 50% of crosslinked dimers should survive cleavage of half the cohesin rings. In contrast, the double ring model predicts only 25% because cleavage of just one ring is sufficient to destroy dimers held together by intertwined rings. We isolated minichromosome dimers from cleavable and non-cleavable haploids, from a 1:1 mixture of the two and from heterozygous diploids. These were crosslinked, treated either with TEV protease or a non-catalytic TEV mutant, denatured with SDS and run on an agarose gel. The fraction of DNAs dimerized by crosslinking was measured by scanning Southern blots. This showed that about 50% of cohesed minichromosomes survived TEV treatment when isolated from heterozygous diploids as well as the 1:1 mixture of haploids (Fig. 4c). These data fit the single but not the double ring model. We note that the latter model also predicts that a sizeable fraction of cross-linked dimers (species 7 and 8) from heterozygous diploids should be converted by TEV cleavage to supercoiled monomers associated with cohesin (species 6), which is not observed (Fig. 4c).

### Discussion

Our crosslinking experiments are consistent with the notion that sister minichromosome DNAs are entrapped by a single monomeric ring. Notably, they exclude the possibility that the connection between sister DNAs is mediated by non-topological interactions between cohesin complexes associated with each sister<sup>8</sup>. Given the specificity

of the crosslinking by bBBR and BMOE, there is no reason to suppose that putative interactions between cohesin rings will have been cross-linked in our experiments. Double-ring models proposing topological cohesin–cohesin interactions or a gigantic ring formed by two cohesin complexes are difficult to reconcile with the findings that the fraction of DNAs dimerized is almost identical to the fraction of cohesin rings circularized and not to the square of this fraction, and that cleavage of half the cohesin rings reduces dimers to 50% and not 25%.

Our conclusion that sister chromatin fibres of dimeric minichromosomes are threaded through cohesin rings provides a simple and potentially adequate mechanism to explain the ability of cohesin to hold sister chromatids together. Cohesin could therefore be considered a ‘concatenase’. It will be important to address whether it uses the same mechanism at loci farther away from core centromeres, whether it sometimes traps individual chromatin fibres, and if so whether it is capable of forming chromatin loops. We detected rare instances where the individual DNA trapping occurred on the minichromosomes, namely monomeric DNAs that on cohesin circularization are retarded in their electrophoretic mobility (species 6) in a



**Figure 4 | Minichromosomes are cohesed by single cohesin rings.**

Protein–protein crosslinking efficiencies by bBBR and BMOE were measured by spiking dimer fractions (K14856) with purified double cysteine mutants of Smc1–Smc3 hinge or Smc1–Scc1–C complex preparations. **a**, After crosslinking, half the reactions were separated by SDS–PAGE and band intensities were measured after SYPRO Ruby staining. **b**, The remainder of the reactions was run on an agarose gel and the fractions of cohesed minichromosomes (7 + 8/total) were quantified after Southern blotting. **c**, Minichromosome dimer fractions from haploid strains expressing TEV-cleavable Smc3–Scc1 (K14856), from haploids expressing Smc3–Scc1 resistant to TEV cleavage (K15089), from a 1:1 mixture of the two haploids, or from a diploid strain containing one TEV-cleavable and one TEV-resistant *SMC3–SCC1* allele (K15267) were crosslinked with bBBR, treated with wild-type or a non-catalytic (nc; C151A) mutant of TEV protease, denatured and separated by agarose gel electrophoresis. Peaks corresponding to cohesed supercoiled minichromosome dimers (7) were integrated and the peak area after incubation with active versus non-catalytic TEV protease was compared. Assuming the number of cohesin cross-bridges follows a Poisson distribution, we would expect 53% or 27% of crosslinked dimers surviving TEV treatment for single or double ring models, respectively. The observed value is 54% ± 6% ( $n = 3$ ; error  $\pm$  s.d.); cleav., cleavable; res., resistant.

manner that is destroyed by ring cleavage. Cohesin is known to associate with chromatin before DNA replication, when it cannot be involved in holding sisters together. Moreover, it can associate with replicated chromosomes in a way that does not lead to cohesion between sisters<sup>12,14</sup>. We suggest that cohesin frequently does trap individual chromatin fibres and that its activity in post-mitotic cells in Metazoa might involve this type of action<sup>15,16</sup>. If we are correct in concluding that cohesin is a novel type of concatenase, then it is not implausible to imagine that other SMC–kleisin complexes such as condensin and its bacterial equivalent have related activities. Indeed, the deep evolutionary roots of these types of complexes suggest that the ability to concatenate DNA may have been an activity without which DNA genomes could not have evolved.

## METHODS SUMMARY

To establish cysteine crosslinking reactions, the Smc1 and Smc3 hinge domains were co-expressed in *Escherichia coli* and purified by means of a C-terminal His<sub>6</sub> tag on Smc1 and gel filtration. The Smc1 head–Scc1–C complex was co-expressed in insect cells using the baculovirus system and purified as described<sup>12</sup>. Proteins were incubated with DMSO or final concentrations of 200  $\mu$ M bBBR or 1 mM BMOE before SDS–PAGE and Coomassie staining. Minichromosomes were isolated as described<sup>7</sup> following a slightly modified protocol. Gradient dimer fractions were dialysed against reaction buffer and incubated for 10 min with DMSO, 200  $\mu$ M bBBR or 1 mM BMOE. Reactions were quenched by addition of DTT to 10 mM and proteins were denatured by 4 min incubation at 65 °C in 1% SDS before agarose gel electrophoresis and Southern blotting. Where indicated, quenched crosslinking reactions were incubated for 1 h at 30 °C with TEV protease at 0.2 mg ml<sup>−1</sup>. To measure crosslinking efficiencies in minichromosome fractions, purified protein dimers were added at a final concentration of 3  $\mu$ M to ensure saturated association of the nanomolar affinity interactions between the dimer subunits.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 5 December 2007; accepted 15 May 2008.**

**Published online 2 July 2008.**

1. Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74**, 595–648 (2005).

2. Hirano, T. SMC proteins and chromosome mechanics: from bacteria to humans. *Phil. Trans. R. Soc. B* **360**, 507–514 (2005).
3. Uhlmann, F. *et al.* Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103**, 375–386 (2000).
4. Haering, C. H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–788 (2002).
5. Hirano, M. & Hirano, T. Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J.* **21**, 5733–5744 (2002).
6. Gruber, S., Haering, C. H. & Nasmyth, K. Chromosomal cohesin forms a ring. *Cell* **112**, 765–777 (2003).
7. Ivanov, D. & Nasmyth, K. A physical assay for sister chromatid cohesion *in vitro*. *Mol. Cell* **27**, 300–310 (2007).
8. Milutinovich, M. & Koshland, D. E. Molecular biology. SMC complexes–wrapped up in controversy. *Science* **300**, 1101–1102 (2003).
9. Kim, J. S. & Raines, R. T. Dibromobimane as a fluorescent crosslinking reagent. *Anal. Biochem.* **225**, 174–176 (1995).
10. Dhar, G., Sanders, E. R. & Johnson, R. C. Architecture of the *hin* synaptic complex during recombination: the recombinase subunits translocate with the DNA strands. *Cell* **119**, 33–45 (2004).
11. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **Supplement 15**, 5.6.1–5.6.30, doi:10.1002/0471250953.bi0506s15 (2006).
12. Haering, C. H. *et al.* Structure and stability of cohesin's Smc1–kleisin interaction. *Mol. Cell* **15**, 951–964 (2004).
13. Gruber, S. *et al.* Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* **127**, 523–537 (2006).
14. Bausch, C. *et al.* Transcription alters chromosomal locations of cohesin in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 8522–8532 (2007).
15. Schuldiner, O. *et al.* piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Dev. Cell* **14**, 227–238 (2008).
16. Pauli, A. *et al.* Cell-type-specific TEV protease cleavage reveals cohesin functions in *Drosophila* neurons. *Dev. Cell* **14**, 239–251 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Gruber for help with the design of cysteine mutations, and P. Fowler, D. Ivanov, V. Katis and all members of the Nasmyth laboratory for advice and discussions. This work was supported by the Medical Research Council, the Wellcome Trust and Cancer Research UK.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.N. ([kim.nasmyth@bioch.ox.ac.uk](mailto:kim.nasmyth@bioch.ox.ac.uk)).



## METHODS

**Model of the *Saccharomyces cerevisiae* Smc1–Smc3 hinge structure.** A structure model of the yeast Smc1–Smc3 hinge structure was created with the Modeller program<sup>11</sup> using an alignment of *S. cerevisiae* Smc1 amino acid residues 488–690 and Smc3 amino acid residues 496–699 with residues 475–679 of the *T. maritima* SMC protein and the coordinates of PDB accession code 1GXL.

**Expression, purification and crosslinking of cohesin subunit domains.** Sequences encoding amino acids 494–705 of the *S. cerevisiae* Smc3 hinge domain followed by an internal ribosome binding site and sequences encoding amino acids 486–696 of the *S. cerevisiae* Smc1 hinge domain fused to a C-terminal His<sub>6</sub> tag were cloned by PCR into the pET28 expression vector. Cysteine mutations were introduced by overlap extension PCR. The Smc1–Smc3 hinge domains were co-expressed in *E. coli* strain BL21(DE3)-RIPL (Stratagene) at 20 °C for 5 h after induction with 0.25 mM IPTG. Cells were lysed in 50 mM NaPi pH 8.0, 300 mM NaCl containing Complete EDTA free protease inhibitor mix (Roche) and the complex was purified using Ni<sup>2+</sup>-chelating affinity chromatography followed by gel filtration on a Superdex 200pg 26/60 column (GE Healthcare) in TEN buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM Na<sub>3</sub>N<sub>3</sub>) plus 100 mM NaCl and 2 mM DTT. The Smc1 head domain bound to Scc1-C was expressed in insect cells using the baculovirus system and purified as described<sup>12</sup>.

Purified proteins were re-buffered into reaction buffer (25 mM NaPi pH 7.4, 50 mM NaCl, 10 mM MgSO<sub>4</sub>, 0.25% Triton X-100) using a Superdex G-25 column, adjusted to 0.5 mg ml<sup>-1</sup> and mixed quickly into one-twenty-fifth volume of DMSO, 5 mM bBBr (Sigma) or 25 mM BMOE (Pierce). Both crosslinkers were dissolved in DMSO just before use. After 10 min incubation at 4 °C, sample loading buffer containing β-mercaptoethanol was added, the samples were heated for 3 min at 90 °C and run on SDS-PAGE followed by Coomassie blue staining. Crosslinking reached a maximum after a few minutes at 4 °C (data not shown).

**Yeast strains.** All strains are derived from W303. Genotypes are listed in Supplementary Table 1.

**Minichromosome preparation and crosslinking.** Yeast strains containing the 2.3 kb minichromosome were grown, arrested in nocodazole and lysed by spheroplasting as described<sup>7</sup>, with the exception that sodium citrate and sodium sulphate in the lysis buffer were replaced by 300 mM NaCl to increase minichromosome yield. Extracts were loaded onto an SW41 10–30% sucrose gradient in 25 mM HEPES-KOH pH 8.0, 50 mM KCl, 10 mM MgSO<sub>4</sub>, 0.25% Triton X-100, 1 mM DTT, 1 mM PMSF. Gradients were run for 15 h at 18,000 r.p.m. and fractionated. Fraction aliquots were separated on a 1% agarose gel containing 0.5 μg ml<sup>-1</sup> ethidium bromide as described<sup>7</sup>. Gels were transferred under alkaline conditions by capillary blotting onto Immobilon-NY<sup>+</sup> membrane (Millipore). The blots were hybridized with a <sup>32</sup>P-labelled probe for the 2.3 kb minichromosome sequence, exposed to imaging plates, scanned on an FLA-7000 image analyser (Fujifilm) and quantified using ImageQuant.

Minichromosome monomer or dimer peak fractions (~300 μl) were dialysed for 4 h against 500 ml reaction buffer at 4 °C in a Float-a-lyzer (SpectraPor) with a 100 kDa molecular mass cutoff. The dialysis buffer was replaced three times and then 24 μl of dialysed fraction was mixed quickly into 1 μl DMSO, 5 mM bBBr or 25 mM BMOE (both freshly dissolved in DMSO) and incubated at 4 °C for 10 min. Final concentrations of 200 μM bBBr or 1 mM BMOE were optimal for crosslinking (Supplementary Fig. 7). Bis-maleimide-based crosslinkers with longer spacers than BMOE, like BMB or BMH (Pierce), could be used with similar efficiency (data not shown). The reaction was quenched by the addition of 1.25 μl 210 mM DTT. For TEV cleavage, 24 μl of the quenched crosslinking reaction was mixed with 1 μl 5 mg ml<sup>-1</sup> wild-type or C151A mutant TEV protease in TEV buffer (TEN buffer plus 50 mM NaCl and 2 mM DTT) or TEV buffer only and incubated at 30 °C for 1 h. Protein was denatured for 4 min at 65 °C after the addition of 2.8 μl 10% SDS. The denatured samples were mixed with 3 μl 80% sucrose containing 0.02% bromophenol blue and 20–25 μl of the mixture were loaded onto a 0.8% agarose gel containing 0.5 μg ml<sup>-1</sup> ethidium bromide. Gels were run at 4 °C for 14 h at 1.4 V cm<sup>-1</sup> and blotted and hybridized as before.

For two-dimensional gels, lanes from the first dimension agarose gels were cut out and placed at the top of a second 0.8% agarose gel (20 × 20 cm) containing 0.5 μg ml<sup>-1</sup> ethidium bromide, leaving an approximately 5-mm-wide slot between the lane and the gel. The slot was filled with 60 °C warm 0.8% agarose in TAE. Proteinase K was dissolved in TAE and mixed with the pre-warmed agarose solution to a final concentration of 0.2 mg ml<sup>-1</sup> just before casting.

Second dimension gels were run at 4 °C for 8 h at 2 V cm<sup>-1</sup>, blotted and hybridized as before.

For nicking minichromosome DNA, 200 μl crosslinked samples were first dialysed against 500 ml reaction buffer plus 2 M KCl and 1 mM DTT at 4 °C for 4 h to remove nucleosomes, then dialysed for 2 h against reaction buffer and 1 mM DTT to remove salt and finally dialysed for 2 h against reaction buffer with 20% sucrose and 1 mM DTT to re-concentrate the samples. Dialysis buffers were replaced every hour. One microlitre of Nb.BsrDI (10 U μl<sup>-1</sup>) was added to 27 μl of sample followed by 10 min incubation at 50 °C, addition of SDS to 1% and denaturation as above.

**Preparation of concatenated minichromosomes from a *top2* strain.** Strain K15029 was grown at 23 °C to log phase in yeast synthetic drop-out medium without tryptophan (–TRP) plus 2% raffinose. Cultures were diluted to A<sub>600 nm</sub> = 0.15 and α-factor was added to a final concentration of 2 μg ml<sup>-1</sup>. Additional α-factor was added to 1.5 μg ml<sup>-1</sup> each after two 40 min intervals. Cells were collected by centrifugation and α-factor was removed by washing with four culture volume yeast extract peptone (YEP) plus 2% raffinose (YEPR) at 4 °C. Cells were resuspended in 35.5 °C warm YEPR containing 10 μg ml<sup>-1</sup> nocodazole and grown for 1.5 h at 35.5 °C. Cells were collected and genomic DNA was prepared by spheroplasting as described above, with the exception that all steps including lysis were performed at 37 °C using buffers pre-warmed to 35.5 °C. Cleared lysates were loaded on pre-chilled 10%–30% sucrose gradients and processed as described above.

**Centromere loop-out.** Cells were inoculated into YEPR from an overnight culture in –TRP plus 2% raffinose and grown at 30 °C to A<sub>600 nm</sub> = 0.6. Galactose or glucose was added to a final concentration of 2%. The 30 min cultures were then diluted to A<sub>600 nm</sub> = 0.15 arrest with α-factor as described above. Cells were collected by centrifugation, washed with four culture volumes YEP plus 2% glucose (YEPD), resuspended in YEPD containing 10 μg ml<sup>-1</sup> nocodazole and grown for 1.5 h at 30 °C. Genomic DNA was prepared as described in Supplementary Information and digested with EcoRI. Cleavage fragments were resolved on a 0.8% agarose gel, Southern blotted and probed for the *TRP1* gene.

**Genomic DNA preparation.** Yeast cells from 25–30 ml asynchronous culture were resuspended in 200 μl SCE (1 M sorbitol, 0.1 M sodium citrate pH 7.0, 60 mM EDTA) buffer plus 0.1 M β-mercaptoethanol and 1 mg ml<sup>-1</sup> zymolyase T-100 and incubated for 1 h at 37 °C with occasional shaking. Spheroplasts were lysed by addition of 200 μl SDS lysis buffer (2% SDS, 0.1 M Tris-HCl pH 9.0, 50 mM EDTA) and incubated for 5 min at 65 °C followed by addition of 200 μl 5 M potassium acetate and centrifugation at 16,000g after 20 min incubation on ice. The supernatant (450 μl) was mixed with 1 ml isopropanol and 200 μl 5 M ammonium acetate, and precipitated genomic DNA was pelleted by 30 s centrifugation at 2,300g. All excess liquid was removed and the pellet was dissolved by 1 h incubation in 90 μl TE at 37 °C. Ten microlitre 5 M ammonium acetate and 200 μl isopropanol were added and genomic DNA was pelleted as above. The DNA pellet was washed with 1 ml 70% ethanol and dissolved in 50 μl TE containing 0.2 mg ml<sup>-1</sup> RNase A. For restriction digests, 5 μl were used in a 20 μl reaction.

**Purification of catalytically active and inactive TEV protease.** Wild-type and catalytically inactive mutant (C151A) TEV protease<sup>17</sup> fused to an N-terminal His<sub>6</sub>-tag were expressed from the pET9d vector in *E. coli* strain BL21(DE3)-RIPL (Stratagene) after induction with 0.25 mM IPTG for 6 h at 22 °C. Cells were lysed in 50 mM Tris-HCl pH 8.0, 1 mM β-mercaptoethanol, 1 mM PMSF. TEV protease was purified from the clarified extract using Ni<sup>2+</sup>-chelating affinity chromatography and gel filtration on a Superdex 200 pg 26/60 column (GE Healthcare) in TEN buffer plus 50 mM NaCl and 2 mM DTT. TEV protease eluted as a single band at the expected delay volume and was concentrated to 5 mg ml<sup>-1</sup> by ultrafiltration.

**Testing cleavage of the Smc3–Scc1 linker.** Protein extracts from 50 ml asynchronous cultures were prepared by glass bead lysis and the Smc3–Scc1 fusion protein was immunoprecipitated by its C-terminal HA<sub>6</sub>-epitope tag with 16B12 antibody as described previously<sup>6</sup>. Immunoprecipitation beads were split and incubated in TEV buffer containing 0.1 mg ml<sup>-1</sup> TEV protease or TEV buffer only for 2 h at 16 °C before addition of loading buffer, SDS-PAGE and western blotting. The blot was probed with 3F10 antibody (Roche) against the HA<sub>6</sub> epitope and exposed to film.

- Hwang, D. C. *et al.* Characterization of active-site residues of the Nla protease from tobacco vein mottling virus. *Mol. Cell* 10, 505–511 (2000).

## LETTERS

# Strong magnetic fields in normal galaxies at high redshift

Martin L. Bernet<sup>1</sup>, Francesco Miniati<sup>1</sup>, Simon J. Lilly<sup>1</sup>, Philipp P. Kronberg<sup>2,3</sup> & Miroslava Dessauges-Zavadsky<sup>4</sup>

The origin and growth of magnetic fields in galaxies is still something of an enigma<sup>1</sup>. It is generally assumed that seed fields are amplified over time through the dynamo effect<sup>2–5</sup>, but there are few constraints on the timescale. It was recently demonstrated that field strengths as traced by rotation measures of distant (and hence ancient) quasars are comparable to those seen today<sup>6</sup>, but it was unclear whether the high fields were in the unusual environments of the quasars themselves or distributed along the lines of sight. Here we report high-resolution spectra that demonstrate that the quasars with strong Mg II absorption lines are unambiguously associated with larger rotation measures. Because Mg II absorption occurs in the haloes of normal galaxies<sup>7–11</sup> along the sightlines to the quasars, this association requires that organized fields of surprisingly high strengths are associated with normal galaxies when the Universe was only about one-third of its present age.

In a recent study<sup>6,12</sup> we used a large sample of extragalactic radio sources to investigate the redshift evolution of the Faraday rotation measure (FRM) of polarized quasars up to  $z \approx 3$ . In agreement with the suggestions of some earlier studies with more limited data<sup>13–16</sup>, we found that the dispersion in the FRM distribution of quasars increases at higher redshifts. This is a non-trivial finding, owing to the strong  $(1+z)^{-2}$  dilution effect, which works to have the opposite outcome. The physical interpretation of this result remained uncertain, however; in particular, there was no way of knowing whether the increase in FRM observed at higher redshifts was associated with an evolution of the unusual environments of the very luminous radio quasars (an ‘intrinsic’ origin), or with the presence of magnetic fields in more typical environments distributed along the sightlines to these cosmologically distant sources (an ‘intervening’ origin).

To test between these possibilities, and thereby resolve a long-standing debate<sup>13–16</sup> about the cause of redshift evolution of the FRM in a clear and decisive way, we have now obtained high-resolution spectra for 76 quasars, selected from our full FRM sample<sup>6</sup>. The spectra (see Supplementary Information) were taken using UVES<sup>17</sup>, the ultraviolet and visual echelle spectrograph on the European Southern Observatory’s 8-m Very Large Telescope. This enabled us to conduct a survey of the intervening material along each individual sightline. The quasars were selected for spectroscopy by requiring that their redshifts lie in the range  $0.6 < z < 2.0$ , spanning the range where the FRM distribution is seen to broaden<sup>6</sup>. Only quasars at Galactic latitudes  $|b| > 30^\circ$  were observed, to minimize problems of Galactic foreground FRM contamination. To ensure adequate signal-to-noise ratios in the optical spectra, the quasars were all selected to be brighter than  $m_V \approx 19$ . No selection was made on the basis of the observed FRMs of the quasars.

Mg II absorption systems in the optical spectra were identified through the double absorption feature at rest-frame wavelengths 2,796.35 and 2,803.53 Å. We catalogued Mg II absorption systems

with rest-frame equivalent widths greater than  $EW_0 > 0.3$  Å in the 2,796 Å line, where our sample should be complete. Because of reduced signal-to-noise ratios in the ultraviolet, we discarded the spectra below 3,760 Å and, thus, any Mg II systems lying below redshift  $z = 0.35$  (few would be expected). To preserve the objectivity of the experiment, the detection of Mg II systems was carried out blind with regard to the FRMs of the quasars.

To provide the cleanest possible test of the intervenor hypothesis, we removed two quasars from the sample that exhibit Mg II absorption at the same redshift as the quasar itself, and discarded three other quasars in which the optical and radio sightlines are separated by more than 7 arcsec ( $\geq 50$  kpc at the redshift of the Mg II absorption). However, there is no significant change in our results if these deletions are not made. Our final sample thus consists of 71 quasar spectra which together cover a total redshift path interval of  $\Delta z = 52.1$ . This contains 36 strong Mg II absorbers with  $EW_0 > 0.3$  Å and redshifts up to  $z = 1.944$  (26 sightlines containing single Mg II systems and five sightlines containing two Mg II systems).

Figures 1 and 2 clearly show that the FRM distribution for sightlines that pass through a strong Mg II system is significantly broader than that for quasars in which absorption is absent. A two-sided Kolmogorov–Smirnov test rejects the possibility that the distributions with  $N_{\text{Mg II}} = 0$ ,  $N_{\text{Mg II}} > 0$  and  $N_{\text{Mg II}} = 0$ ,  $N_{\text{Mg II}} = 2$  (where  $N_{\text{Mg II}}$  denotes the number of observed Mg II systems) respectively come from the same parent distribution at the 92.2% and 99.93% levels, or 96.1% for the sightlines with quasar redshift  $z > 1$ . We note that the greatest deviation occurs at an FRM value of  $\sim 40 \text{ rad m}^{-2}$ , which corresponds to the median FRM of the sample, showing that the result is produced by the bulk of the data and not by a few extreme outliers. This result is obtained without any subtraction of a Galactic foreground model, but remains essentially unchanged if the individual foreground corrections are applied, as in ref. 6. Statistical completeness and/or sensitivity to the Galactic foreground subtraction have been major challenges in previous analyses<sup>16</sup>.

Because the probability of intersecting a Mg II absorber along the sightline increases strongly with the redshift of the quasar, the correlation of  $|\text{FRM}|$  (the absolute value of the FRM) with Mg II absorption systems in Figs 1 and 2 might be thought to arise because of an underlying correlation between the intrinsic component of the quasar FRM and the quasar redshift, due, for example, to an evolutionary effect in the quasars themselves. However, the correlation between  $|\text{FRM}|$  and the redshift of the quasar,  $z$ , is evidently much weaker than that between  $|\text{FRM}|$  and  $N_{\text{Mg II}}$ : the former has a Kendall’s  $\tau$  statistic of  $\tau = 0.11$  and a chance probability of 0.18, whereas the latter has  $\tau = 0.25$  and the chance probability of only 0.008. A more transparent test is to simply compare the  $|\text{FRM}|$  values of those quasars that have Mg II absorbers with the median  $|\text{FRM}|$  of the seven quasars without absorbers that lie closest in redshift. If Mg II absorption does not

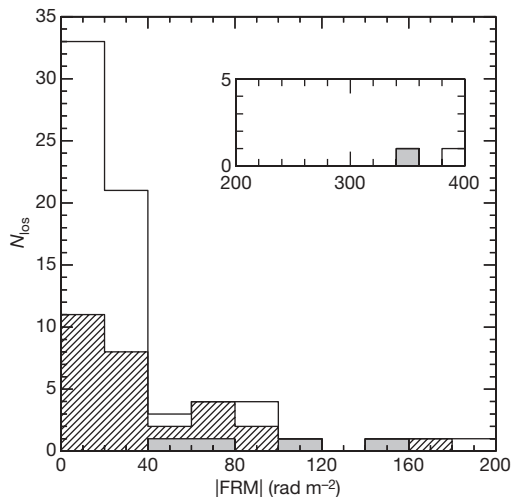
<sup>1</sup>Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 16, CH-8093 Zürich, Switzerland. <sup>2</sup>Los Alamos National Laboratory, IGPP, PO Box 1663, Los Alamos, New Mexico 87545, USA. <sup>3</sup>Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario M5S 1A7, Canada. <sup>4</sup>Observatoire de Genève, Université de Genève, 51 Chemin des Maillettes, CH-1290 Sauverny, Switzerland.



contribute to the FRM, then the  $|\text{FRM}|$  values of absorber quasars should be randomly distributed above and below this median. We find that the  $|\text{FRM}|$  values of 20 of the 31 absorber quasars lie above the local median  $|\text{FRM}|$  value of non-absorber quasars, a result which, from simple binomial statistics, has a chance probability of occurring of 7%. We conclude that the observed correlation between  $|\text{FRM}|$  and  $N_{\text{Mg II}}$  is a real physical effect.

The Mg II absorption systems are known to occur in the haloes of normal galaxies<sup>7,8</sup> characterized by a wide range of colours and optical luminosities ( $0.1L^* - 3.0L^*$ , where  $L^* \approx 2 \times 10^{10} L_{\odot} - 3 \times 10^{10} L_{\odot}$  is a characteristic luminosity of the galaxies that dominate the light output, and  $L_{\odot}$  is the solar luminosity), with a covering factor close to unity for impact parameters  $D \leq 50$  kpc (ref. 9). This means that it is possible to find a parent galaxy for virtually all Mg II systems within this radius, and that few sightlines pass this close to a galaxy and yet do not produce Mg II absorption. More recent investigations detected Mg II absorption even out to distances as large as  $\sim 100$  kpc<sup>10,11</sup>, but with a smaller covering factor<sup>11</sup>.

With the new information about the individual sightlines from the optical spectra, we can now obtain an improved determination of the FRM contribution of the intervening systems with respect to that in ref. 6 (see Supplementary Information). We replace the statistical description of the occurrence of Mg II systems with the actual information for each sightline. We keep the intrinsic lorentzian component to a constant (rest-frame)  $F_{\text{int}}/2 = 7_{-2}^{+6} \text{ rad m}^{-2}$ , as obtained in our previous analysis<sup>6</sup>, because in that work the many sources at low redshifts  $z < 0.6$  allowed for the optimal determination of this parameter. Our new calculation yields  $\sigma_{\text{Mg II}} = 140_{-50}^{+80} \text{ rad m}^{-2}$  and  $\sigma_{\text{noise}} = 20_{-4}^{+10} \text{ rad m}^{-2}$  ( $\pm 1$  s.d. uncertainties) for the dispersions of, respectively, the rest-frame FRM distribution of Mg II systems and the Galactic noise. At redshift  $z \approx 0.9 - 1.3$ , this leads to an additional FRM component in the observer's frame of about  $40 \text{ rad m}^{-2}$ , as seen in Figs 1 and 2. These results indicate that the largest contribution to the observed Faraday rotation of typical quasars may be coming from the intervening galaxies, although undoubtedly some individual quasars will exhibit high intrinsic values of FRM<sup>18</sup>.



**Figure 1 | FRM distributions for different numbers of strong Mg II absorption lines.** The FRM histogram of the total sample consisting of 71 sightlines to quasars with redshifts in the range  $z = 0.6 - 2.0$  is shown in the background ( $N_{\text{Mg II}} \geq 0$ ; unshaded). Plotted over this are the histogram of 31 sightlines with one or two strong Mg II absorption line systems ( $N_{\text{Mg II}} \geq 1$ ; hatched) and the histogram of five sightlines with two absorption line systems ( $N_{\text{Mg II}} = 2$ ; grey). The inset has the same axes as the main panel and shows the high- $|\text{FRM}|$  data;  $N_{\text{los}}$ , number of lines of sight. Using Kendall's  $\tau$  statistic, we determine that  $|\text{FRM}|$  and the number of detected Mg II systems are correlated at a significance level of 99.12%. A typical FRM measurement error is around  $3 \text{ rad m}^{-2}$ .

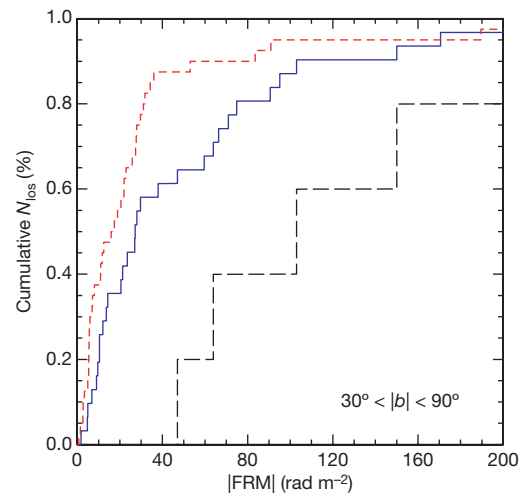
The value of  $\sigma_{\text{Mg II}}$  obtained above for the intervenors is interesting, as it is comparable to the dispersion characterizing the FRM data of known galactic systems at low redshift, for example the disk of the Milky Way<sup>19</sup>, which has  $|\text{FRM}|$  values of up to a few hundred  $\text{rad m}^{-2}$ , the Large Magellanic Cloud<sup>20</sup> and a spiral galaxy with an absorbing system at  $z = 0.395$  in front of the quasar PKS 1229-021 (ref. 21).

For an order-of-magnitude estimate of the magnetic field strength,  $B$ , in the intervening systems, we can re-write the FRM definition (see Supplementary Information) in the rest frame of the Mg II system as  $\text{FRM} \approx 1.5 \times 10^{-19} N_e B$ , where FRM is measured in  $\text{rad m}^{-2}$ ;  $N_e$ , the column density of free electrons, is measured in  $\text{cm}^{-2}$ ; and  $B$  is measured in  $\mu\text{G}$ . Typically, Mg II systems with  $0.3 \text{ \AA} < \text{EW}_0 < 0.6 \text{ \AA}$  are found to be associated with galaxies with a neutral-hydrogen column density of  $N(\text{H I}) \approx 10^{19} \text{ cm}^{-2}$  (ref. 22) and a hydrogen ionization fraction of  $\bar{x} \approx 0.90$  (refs 23, 24), yielding  $N_e \approx \bar{x}/(1 - \bar{x})N(\text{H I})$ , so that:

$$B \approx 10 \left( \frac{\sigma_{\text{Mg II}}}{140 \text{ rad m}^{-2}} \right) \left( \frac{N_e}{9 \times 10^{19} \text{ cm}^{-2}} \right)^{-1} \mu\text{G}$$

This order-of-magnitude estimate ignores any field reversals within the system, which would boost the required magnetic fields by the square root of the number of such reversals.

The optical and radio sightlines to quasars showing Mg II absorption are aligned to within 2 arcsec (25 kpc at  $z = 1$ ) of each other in over 90% of the objects. Most of the radio sources are also compact, with angular sizes of about 1 arcsec, that is, covering an area of diameter less than 10 kpc in the redshift range of  $0.6 < z < 2.0$ . Because the impact parameters of Mg II systems generally are as great as 50 kpc (ref. 9), or even 100 kpc (refs 6, 10), measured from the parent galaxies, it is thus likely that our radio sightlines are probing the outer haloes of galaxies in addition to the inner disks. The magnetized, enriched material could have been ejected by galactic winds. However, further observations to locate the parent galaxies of these particular Mg II systems relative to the individual radio sightlines will be required before we can fix the location of the magnetized material within these galaxies.



**Figure 2 | Cumulative FRM distributions for sightlines with and without strong Mg II absorption line systems.** The 31 sightlines with one or two strong Mg II systems ( $N_{\text{Mg II}} \geq 1$ ) correspond to the blue solid line, those with two strong Mg II systems ( $N_{\text{Mg II}} = 2$ ) correspond to the black dashed line and the 40 sightlines without strong Mg II absorption lines ( $N_{\text{Mg II}} = 0$ ) are represented by the red dashed line. Cumulative  $N_{\text{los}}$ , fraction of number of lines of sight for which  $|\text{FRM}|$  is less than the abscissa value. For clarity, only the range  $0 \text{ rad m}^{-2} < |\text{FRM}| < 200 \text{ rad m}^{-2}$  is shown. A typical FRM measurement error is around  $3 \text{ rad m}^{-2}$ . A Kolmogorov-Smirnov test indicates that the FRM distributions for  $N_{\text{Mg II}} = 0$ ,  $N_{\text{Mg II}} > 0$  are different at the 92.2% significance level (96.1% for quasar redshift  $z > 1$ ) and that those for  $N_{\text{Mg II}} = 0$ ,  $N_{\text{Mg II}} = 2$  are different at the 99.93% significance level.

We conclude that magnetic fields in and around completely normal galaxies were, at the time when the Universe was only about one-third ( $z \approx 1.3$ ) of its present age, already of comparable strength to those that are observed in today's galaxies<sup>25,26</sup>. This significantly reduces the number of e-foldings available for the build-up of coherent magnetic fields for alpha-omega dynamo mechanisms in galaxies, setting new, strong and general constraints on the efficiency of those models<sup>3,27</sup>. It also serves as a general reminder of the potential importance of magnetic fields, which is usually completely ignored, in the formation and evolution of cosmic structures in the high-redshift Universe.

Received 23 November 2007; accepted 13 May 2008.

- Rees, M. J. Origin of cosmic magnetic fields. *Astron. Nachr.* **327**, 395–398 (2006).
- Zel'dovich, YaB., Ruzmaikin, A. A. & Sokoloff, D. D. in *Magnetic Fields in Astrophysics* (ed. Roberts, P. H.) 37–44 (Gordon & Breach, Montreux, 1983).
- Kulsrud, R. M. & Zweibel, E. G. On the origin of cosmic magnetic fields. *Rep. Prog. Phys.* **71**, doi:10.1088/0034-4885/71/4/046901 (2008).
- Kulsrud, R. M. A critical review of galactic dynamos. *Annu. Rev. Astron. Astrophys.* **37**, 37–64 (1999).
- Parker, E. N. Fast dynamos, cosmic rays, and the Galactic magnetic field. *Astrophys. J.* **401**, 137–145 (1992).
- Kronberg, P. P. et al. A global probe of cosmic magnetic fields to high redshifts. *Astrophys. J.* **676**, 70–79 (2008).
- Churchill, C. W., Kacprzak, G. G. & Steidel, C. C. in *Probing Galaxies through Quasar Absorption Lines* (Proc. IAU Colloq. C19) (ed. Roberts, P. H.) 24–41 (Cambridge Univ. Press, Cambridge, UK, 2005).
- Nestor, D. B., Turnshek, D. A. & Rao, S. M. Mg II absorption Systems in Sloan Digital Sky Survey QSO spectra. *Astrophys. J.* **628**, 637–654 (2005).
- Steidel, C. C. in *QSO Absorption Lines* (Proc. ESO Workshop, Garching, Germany, 21–24 November 1994) (ed. Meylan, G.) 139–152 (Springer, Berlin, 1995).
- Zibetti, S. et al. Optical properties and spatial distribution of Mg II absorbers from SDSS image stacking. *Astrophys. J.* **658**, 161–184 (2007).
- Kacprzak, G. G., Churchill, C. W., Steidel, C. C. & Murphy, M. T. Halo gas cross sections and covering fractions of MgII absorption selected galaxies. *Astron. J.* **135**, 922–927 (2008).
- Bernet, M. L. *Cosmic Magnetic Fields at High Redshifts*. Diploma thesis, ETH Zürich (2005).
- Kronberg, P. P. & Perry, J. J. Absorption lines, Faraday rotation, and magnetic field estimates for QSO absorption-line clouds. *Astrophys. J.* **263**, 518–532 (1982).
- Welter, G. L., Perry, J. J. & Kronberg, P. P. The rotation measure of QSOs and of intervening clouds – Magnetic fields and column densities. *Astrophys. J.* **279**, 19–39 (1984).
- Watson, A. M. & Perry, J. J. QSO absorption lines and rotation measure. *Mon. Not. R. Astron. Soc.* **248**, 58–73 (1991).
- Oren, A. L. & Wolfe, A. M. A Faraday rotation search for magnetic fields in quasar damped Ly alpha absorption systems. *Astrophys. J.* **445**, 624–641 (1995).
- Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B. & Kotzlowski, H. Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal Observatory. *Proc. SPIE* **4008**, 534–545 (2000).
- Kronberg, P. P., Perry, J. J. & Zukowski, E. L. The 'jet' rotation measure distribution and the optical absorption system near the  $z = 1.953$  quasar 3C191. *Astrophys. J.* **355**, L31–L34 (1990).
- Han, J. L. et al. Pulsar rotation measure and the large-scale structure of the Galactic magnetic field. *Astrophys. J.* **642**, 868–881 (2006).
- Gaensler, B. M. et al. The magnetic field of the Large Magellanic Cloud revealed through Faraday rotation. *Science* **307**, 1610–1612 (2005).
- Kronberg, P. P., Perry, J. J. & Zukowski, E. L. Discovery of extended Faraday rotation compatible with spiral structure in an intervening galaxy at  $z = 0.395$  – New observations of PKS 1229–021. *Astrophys. J.* **387**, 528–535 (1992).
- Rao, S. M., Turnshek, D. A. & Nestor, D. B. Damped Ly $\alpha$  systems at  $z < 1.65$ : The expanded Sloan Digital Sky Survey Hubble Space Telescope sample. *Astrophys. J.* **636**, 610–630 (2006).
- Pérout, C., Dessauges-Zavadsky, M., D'Odorico, S., Kim, T. & McMahon, R. G. A homogenous sample of sub-damped Lyman systems – IV. Global metallicity evolution. *Mon. Not. R. Astron. Soc.* **382**, 177–193 (2007).
- Prochaska, J. X. et al. Supersolar super-Lyman limit systems. *Astrophys. J.* **648**, 97–100 (2006).
- Beck, R., Brandenburg, A., Moss, D., Shukurov, A. & Sokoloff, D. Galactic magnetism: Recent developments and perspectives. *Annu. Rev. Astron. Astrophys.* **34**, 155–206 (1996).
- Beck, R. The role of magnetic fields in spiral galaxies. *Astrophys. Space Sci.* **289**, 293–302 (2004).
- Kronberg, P. P. Extragalactic magnetic fields. *Rep. Prog. Phys.* **57**, 325–382 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Our observations were made on the European Southern Observatory's telescopes at the Paranal Observatory under programme IDs 075.A-0841 and 076.A-0860. M.L.B. acknowledges financial support from the Swiss National Science Foundation, and P.P.K. acknowledges support from the Natural Sciences and Engineering Council of Canada and the US Department of Energy.

**Author Contributions** M.L.B. reduced the UVES spectra, identified the Mg II absorbers in them and carried out the statistical analyses that are presented in the paper. F.M. was the principal investigator on the UVES observational project on the Very Large Telescope and made the observations in Chile. F.M. and S.J.L. oversaw the overall design and execution of the research. P.P.K. derived the FRMs from multifrequency polarization measurements obtained at the NRAO Very Large Array and the Max Planck Institute for Radio Astronomy's Effelsberg 100-m radio telescope, and other radio telescopes. M.D.-Z. advised on the observation and reduction of the UVES data and the identification of Mg II absorbers.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to F.M. (fm@phys.ethz.ch).



# Hydrated silicate minerals on Mars observed by the Mars Reconnaissance Orbiter CRISM instrument

John F. Mustard<sup>1</sup>, S. L. Murchie<sup>2</sup>, S. M. Pelkey<sup>1</sup>, B. L. Ehlmann<sup>1</sup>, R. E. Milliken<sup>3</sup>, J. A. Grant<sup>4</sup>, J.-P. Bibring<sup>5</sup>, F. Poulet<sup>5</sup>, J. Bishop<sup>6</sup>, E. Noe Dobrea<sup>3</sup>, L. Roach<sup>1</sup>, F. Seelos<sup>2</sup>, R. E. Arvidson<sup>7</sup>, S. Wiseman<sup>7</sup>, R. Green<sup>3</sup>, C. Hash<sup>8</sup>, D. Humm<sup>2</sup>, E. Malaret<sup>8</sup>, J. A. McGovern<sup>2</sup>, K. Seelos<sup>2</sup>, T. Clancy<sup>9</sup>, R. Clark<sup>10</sup>, D. Des Marais<sup>6</sup>, N. Izenberg<sup>2</sup>, A. Knudson<sup>7</sup>, Y. Langevin<sup>5</sup>, T. Martin<sup>3</sup>, P. McGuire<sup>7</sup>, R. Morris<sup>11</sup>, M. Robinson<sup>12</sup>, T. Roush<sup>6</sup>, M. Smith<sup>13</sup>, G. Swayze<sup>9</sup>, H. Taylor<sup>2</sup>, T. Titus<sup>14</sup> & M. Wolff<sup>9</sup>

Phyllosilicates, a class of hydrous mineral first definitively identified on Mars by the OMEGA (Observatoire pour la Mineralogie, L'Eau, les Glaces et l'Activité) instrument<sup>1,2</sup>, preserve a record of the interaction of water with rocks on Mars. Global mapping showed that phyllosilicates are widespread but are apparently restricted to ancient terrains and a relatively narrow range of mineralogy (Fe/Mg and Al smectite clays). This was interpreted to indicate that phyllosilicate formation occurred during the Noachian (the earliest geological era of Mars), and that the conditions necessary for phyllosilicate formation (moderate to high pH and high water activity<sup>3</sup>) were specific to surface environments during the earliest era of Mars's history<sup>4</sup>. Here we report results from the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM)<sup>5</sup> of phyllosilicate-rich regions. We expand the diversity of phyllosilicate mineralogy with the identification of kaolinite, chlorite and illite or muscovite, and a new class of hydrated silicate (hydrated silica). We observe diverse Fe/Mg-OH phyllosilicates and find that smectites such as nontronite and saponite are the most common, but chlorites are also present in some locations. Stratigraphic relationships in the Nili Fossae region show olivine-rich materials overlying phyllosilicate-bearing units, indicating the cessation of aqueous alteration before emplacement of the olivine-bearing unit. Hundreds of detections of Fe/Mg phyllosilicate in rims, ejecta and central peaks of craters in the southern highland Noachian cratered terrain indicate excavation of altered crust from depth. We also find phyllosilicate in sedimentary deposits clearly laid by water. These results point to a rich diversity of Noachian environments conducive to habitability.

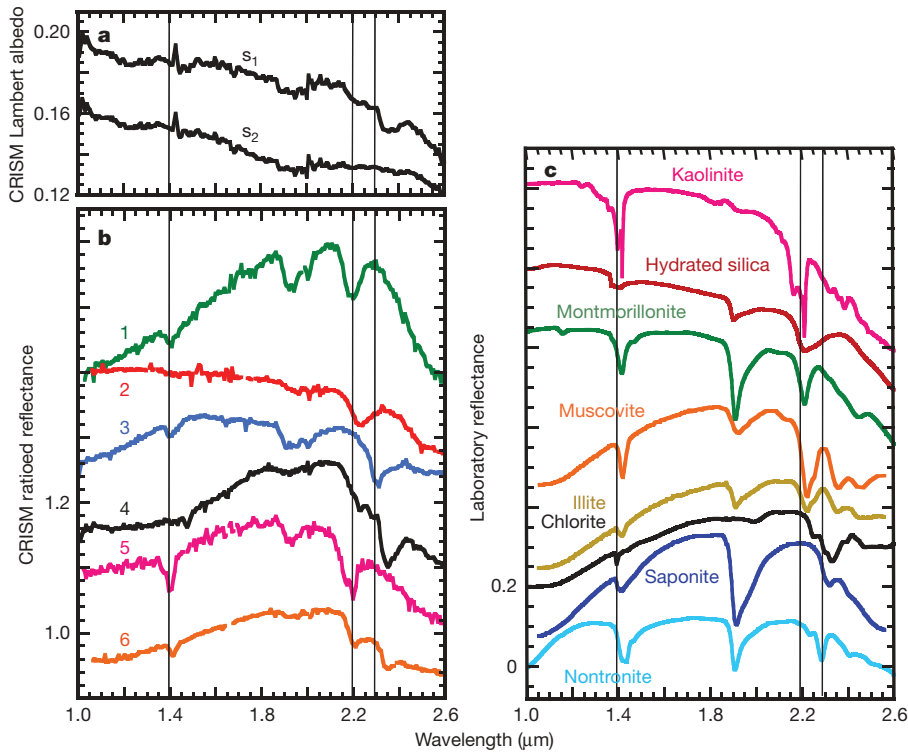
High-spatial-resolution, precision-pointing and nested observations of CRISM<sup>5</sup>, the Context Imager (CTX)<sup>6</sup>, and the High Resolution Imaging Science Experiment (HiRISE)<sup>7</sup> instruments on the Mars Reconnaissance Orbiter (MRO) resolve mineralogical, stratigraphic and geological relationships for phyllosilicate deposits. This combination of instruments permits mineralogical mapping at 18 m per pixel with CRISM linked with metre-scale geomorphology from CTX and HiRISE. We focus here on the stratigraphic setting

of phyllosilicate-bearing rocks in three regions and report the detection of phyllosilicate in sedimentary settings.

We identify two principal classes of mineral in the CRISM data on the basis of observed absorptions: Al phyllosilicates and the more common and spatially dominant Fe/Mg phyllosilicates. The increased spatial and spectral resolutions of CRISM have revealed a diversity of absorption band shapes, positions and combinations indicating variations in phyllosilicate type and composition (Fig. 1; see Methods for processing and identification details). Most spectra show a band at  $\sim 1.4\text{ }\mu\text{m}$  from the overtone of the OH stretch, a strong  $1.9\text{-}\mu\text{m}$  H<sub>2</sub>O band and absorptions near  $2.28\text{--}2.30\text{ }\mu\text{m}$  (for example, spectrum 3 in Fig. 1b), which are consistent with Fe/Mg phyllosilicates having interlayer water, such as the smectite clays nontronite and saponite<sup>8</sup>. Other spectra in Nili Fossae have a dominant absorption near or longward of  $2.34\text{ }\mu\text{m}$ , a weak band or inflection near  $2.25\text{ }\mu\text{m}$ , and a weak or absent  $1.9\text{ }\mu\text{m}$  absorption band (spectrum 4 in Fig. 1b). This combination of absorptions is consistent with high-Fe chlorites<sup>9–11</sup>. Some spectra from Mawrth Vallis show strong  $1.9\text{-}\mu\text{m}$  and  $2.2\text{-}\mu\text{m}$  features and a weak  $1.4\text{-}\mu\text{m}$  band (spectrum 1 in Fig. 1b) diagnostic of montmorillonite, an Al-rich smectite clay, confirming previous detections<sup>2</sup>. K,Al phyllosilicates have also been recognized in a new location, the ejecta of a 50-km crater west of Nili Fossae. These spectra show  $1.4\text{-}\mu\text{m}$ ,  $2.2\text{-}\mu\text{m}$  and  $\sim 2.35\text{-}\mu\text{m}$  absorptions but lack a  $1.9\text{-}\mu\text{m}$  band (spectrum 6 in Fig. 1b). These absorptions are consistent with the presence of mica minerals such as illite and/or muscovite<sup>10</sup>. Additionally, in some small-scale outcrops a distinct doublet absorption with resolved absorptions at  $2.16$  and  $2.2\text{ }\mu\text{m}$ , and a strong  $1.4\text{-}\mu\text{m}$  band is observed (spectrum 5 in Fig. 1b). This combination of bands is diagnostic of kaolinite. Kaolinite was first observed by OMEGA in one small region of the southern highlands<sup>12</sup>, whereas CRISM has observed kaolinite in Nili Fossae and Mawrth Vallis.

A new class of hydrated silicate has been identified (spectrum 2 in Fig. 1b), characterized by a  $2.20\text{--}2.25\text{-}\mu\text{m}$  absorption that is distinct from that observed with Al-OH phyllosilicates such as montmorillonite in that the absorption is broader and centred at longer wavelengths. These regions commonly show  $1.4\text{-}\mu\text{m}$  and  $1.9\text{-}\mu\text{m}$

<sup>1</sup>Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. <sup>2</sup>Johns Hopkins University/Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA. <sup>3</sup>Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 183-301, 4800 Oak Grove Drive, Pasadena, California 91109, USA. <sup>4</sup>Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Independence Avenue at 6th Street SW, Washington, DC 20560, USA. <sup>5</sup>Institut d'Astrophysique Spatiale, Université Paris Sud 11, 91405 Orsay, France. <sup>6</sup>National Aeronautics and Space Administration, Ames Research Center, 515 N. Whisman Road, Mountain View, California 94043, USA. <sup>7</sup>Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA. <sup>8</sup>Applied Coherent Technology, 112 Elden Street Suite K, Herndon, Virginia 22070, USA. <sup>9</sup>Space Science Institute, 4750 Walnut Street, Suite 205, Boulder, Colorado 80301, USA. <sup>10</sup>US Geological Survey, MS 964 Box 25046, Denver Federal Center, Denver, Colorado 80225, USA. <sup>11</sup>ARES Code KR, National Aeronautics and Space Administration, Johnson Space Center, 2101 NASA Parkway, Houston, Texas 77058, USA. <sup>12</sup>School of Earth and Space Exploration, Box 871404, Arizona State University, Tempe, Arizona 85287-1404, USA. <sup>13</sup>National Aeronautics and Space Administration, Goddard Space Flight Center, Code 693.0, Greenbelt, Maryland 20771, USA. <sup>14</sup>US Geological Survey, 2255 N. Gemini Drive, Flagstaff, Arizona, 86001, USA.

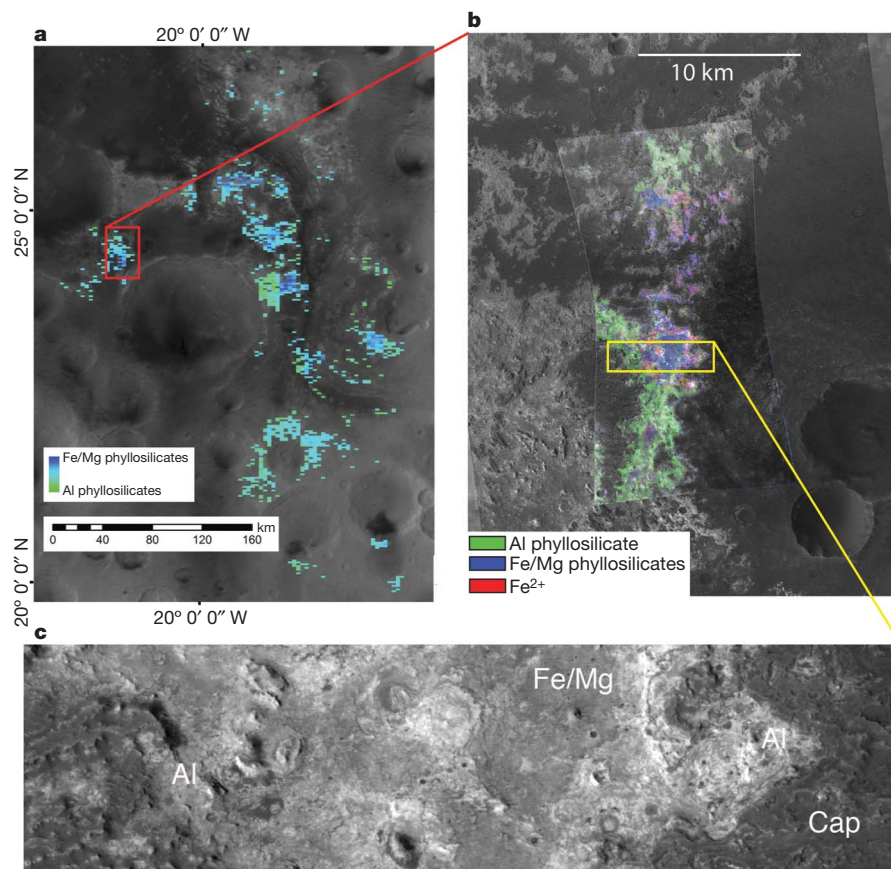


**Figure 1 | CRISM and laboratory reflectance spectra showing hydrated silicate mineral diversity.** **a**, CRISM reflectance spectra from 1.0 to 2.6  $\mu\text{m}$  used in the spectral ratio ( $s_1/s_2$ ) for spectrum 4 in **b**. **b**, Stacked spectral ratios: (1) from Mawrth Vallis, green (HRL0000285A); (2) from the southern highlands, red (FRT00035DB); (3) from Nili Fossae, blue (FRT00003E12); (4) from Nili Fossae, black (FRT00050F2); (5) from Nili Fossae, magenta (FRT00003FB9); (6) from Nili Fossae, orange (HRS0002FC5). Spectra were selected on the basis of strong BD2210 and D2300 parameter values. Vertical lines are placed at 1.4, 2.2 and 2.3  $\mu\text{m}$ . **c**, Stacked library reflectance spectra of minerals with similar absorption features. The hydrated silica is an altered, hydrated volcanic ash.

bands, and this combination of spectral features is consistent with hydrated silica glasses such as opal or volcanic glass.

In the Mawrth Vallis area (Fig. 2), smectites rich in Al and Fe have been observed over a 300 km  $\times$  400 km region but are commonly found along the flanks of the outflow channel and within the

surrounding plateau<sup>13,14</sup>. The units hosting these minerals are typically light-toned outcrops of early-Noachian to mid-Noachian materials that have been exposed by erosion from beneath a dark mantling unit that shows pyroxene absorptions. Some occurrences of phyllosilicate are clearly associated with layered rocks. This leads to



**Figure 2 | Mineral diversity and stratigraphy of Mawrth Vallis.** **a**, MOC map of Mawrth Vallis with OMEGA phyllosilicate mineral indicators<sup>11</sup> for Fe/Mg phyllosilicates (blue) and Al phyllosilicates (green). **b**, CTX image of the red box from **a**, with a superimposed mineral indicator map from CRISM observation HRL000043EC\_07, in which green areas are enriched in Al phyllosilicate, blue areas are enriched in Fe/Mg phyllosilicate, and red areas indicate an  $\text{Fe}^{2+}$  slope. **c**, Subset of the CTX image from the white box in **b**, where the Fe/Mg phyllosilicate (Fe/Mg) occurs as a darker smooth surface. Al phyllosilicates (Al) are within brighter and rougher terrain beneath a capping unit (cap).



the hypothesis that they may have formed by deposition in an aqueous environment, by aqueous alteration of a volcanic ash deposit, or by impact-related alteration<sup>13</sup>.

CRISM observation HRL000043EC falls in the region west of the mouth of Mawrth Vallis (Fig. 2a)<sup>13</sup>. In these light-toned outcrops we find discrete layers of Al and Fe/Mg phyllosilicates. In this region Al phyllosilicates are more prevalent (Fig. 2b); however, elsewhere in Mawrth Vallis Fe/Mg phyllosilicates are typically more abundant. In addition to CRISM spectra showing absorptions diagnostic of montmorillonite as observed previously with OMEGA, we also observe in small outcrops absorptions characteristic of kaolinite and hydrated silica or glass. Sandwiched between the Al and Fe/Mg phyllosilicate layers in CRISM image HRL000043EC is an additional material containing an Fe<sup>2+</sup> slope and a broadened 2.2–2.3- $\mu$ m feature.

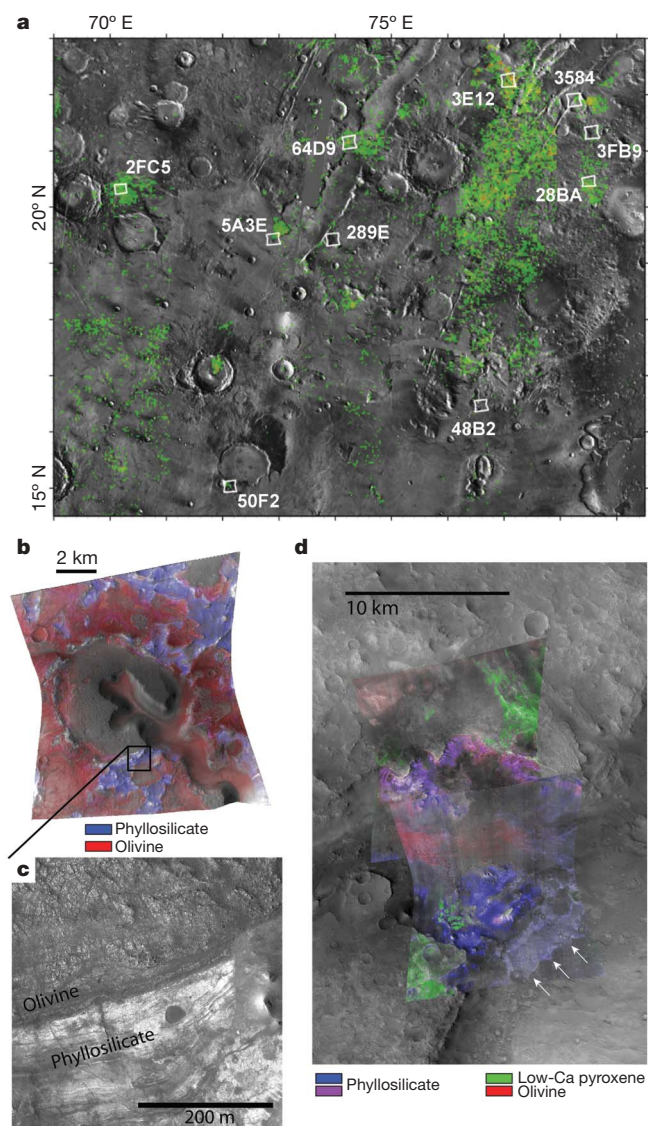
High-resolution Mars Orbiting Camera (MOC) and High Resolution Stereo Camera (HRSC) images<sup>13,14</sup>, as well as new observations acquired by CTX and HiRISE, reveal that the light-toned outcrops in Mawrth Vallis are layered and often show a distinct polygonal texture. The Fe/Mg-phyllosilicate-bearing unit is the lowest stratigraphic level and is overlain by Al-phyllosilicate-bearing strata; both are exposed by erosion from beneath a dark-toned capping unit (Fig. 2c). Despite the close proximity of these phyllosilicate-bearing units, we see no evidence for mixing of the two phyllosilicate signatures at spatial resolutions of  $\sim 40$  m per pixel. This indicates that the minerals occur in spatially distinct stratigraphic units.

The Nili Fossae region shows expansive outcrops of phyllosilicates<sup>2,12</sup> but also extensive spectral signatures of olivine in close proximity<sup>15–19</sup>. Several high-resolution observations across this region (Fig. 3a) show that much of the olivine is present within aeolian bedforms. However, several observations provide definitive evidence for the presence of phyllosilicate-bearing units in direct contact with unaltered olivine-bearing units. In CRISM observation FRT00003E12, stratigraphic relationships are resolved between the lowest, phyllosilicate-bearing bedrock unit and the overlying olivine-bearing bedrock unit (Fig. 3b). The phyllosilicate-rich unit shows polygonal fractures and is recessed on slope-forming units, suggesting that it is more easily eroded than the cap unit. The phyllosilicate unit is capped by a regionally extensive, spectrally neutral, mesa-forming unit. Although most of the cap unit is massive, HiRISE data show the base is tens of metres thick and contains banding and/or layering (Fig. 3c). Coordinated CRISM data show this banded unit is olivine-rich and that it rests directly on top of the phyllosilicate-rich material (Fig. 3c). This sequence of a spectrally neutral cap rock overlying an unaltered olivine-rich unit resting on phyllosilicate-rich rocks is observed in CRISM observations across a distance of 2,000 km from the Nili Fossae region to the southern edge of the Isidis Basin<sup>20</sup>.

CRISM has targeted contacts between phyllosilicate-bearing rocks and Hesperian-aged lavas from Syrtis Major. Some of the observations are of high-standing Noachian-aged massifs embayed by lava (FRT000050F2 and FRT00005A3E) and other observations show phyllosilicate-bearing outcrops beneath volcanic plains (FRT000048B2, and Fig. 3d). In all these observations of volcanic rocks in contact with phyllosilicate-bearing terrains, the volcanic rocks are unaltered and provide no evidence for hydrated silicates. A section of the Nili Fossae trough scarp that is  $\sim 600$  m high exposes rocks that are strongly enriched in phyllosilicates, whereas the rocks capping the plateau are enriched in pyroxene (Fig. 3d).

In the ancient southern highlands, localized phyllosilicate deposits were identified with OMEGA data<sup>19,21</sup>. With the use of a combination of multispectral mapping and targeted observations, CRISM has significantly expanded the number of locations to the thousands of phyllosilicate occurrences (see, for example, Fig. 4a). These data show phyllosilicates associated with craters of many sizes and within ejecta, walls and central peaks. Full-resolution CRISM observation FRT000049BB shows a highlands crater in which Fe/Mg smectite clays can be seen in outcrops at or near the rim of the crater along

its inward facing wall (Fig. 4b). Weaker phyllosilicate detections are visible downslope of these outcrops, indicating physical breakdown and mass wasting. In addition we observe phyllosilicate associated with ejecta exterior to the crater and in mounds and knobs on the crater floor. CRISM observation FRT00003E92 shows Fe/Mg phyllosilicates associated with the central peak of a crater 45 km in diameter (Fig. 4c). This central peak material probably originated from a depth of 4–5 km, indicating alteration of the crust to at least this depth. Elsewhere in the southern highlands we observe phyllosilicates in central peaks that show a distinct 2.2- $\mu$ m absorption indicative of either Al phyllosilicate or hydrated glass (Fig. 1).



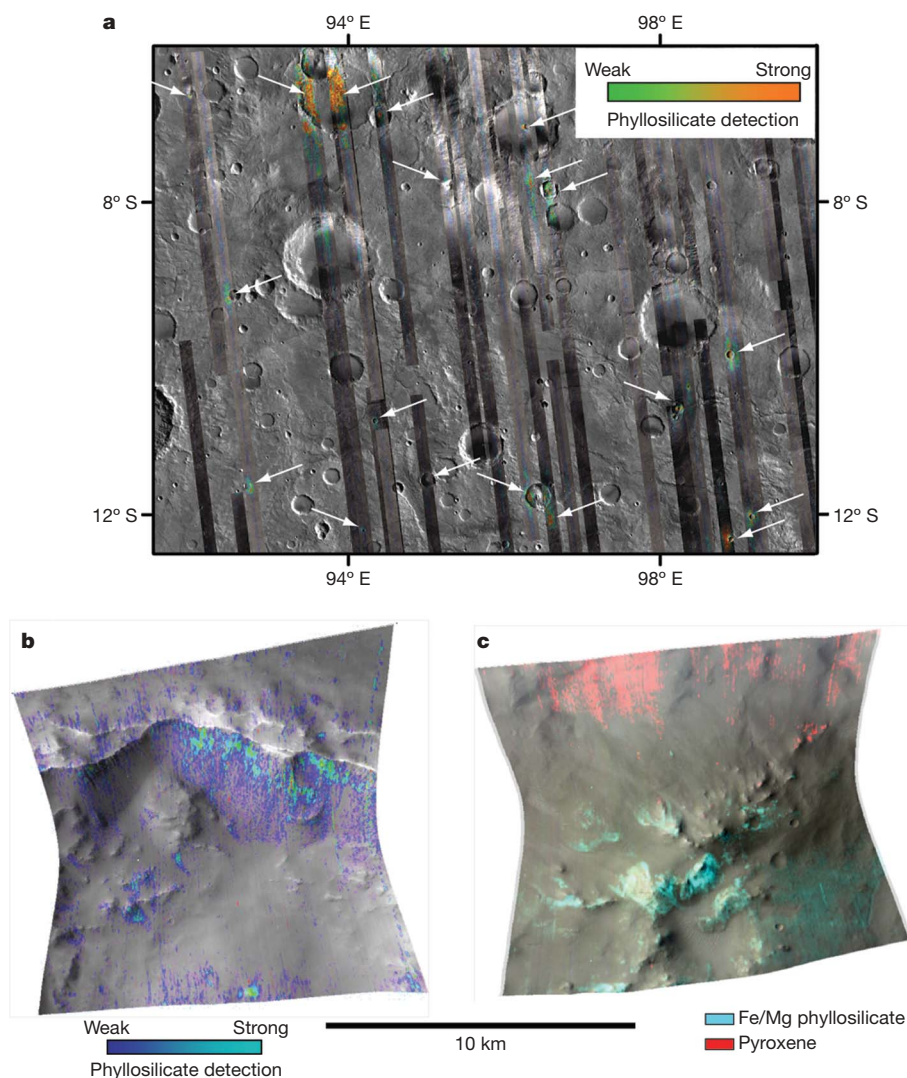
**Figure 3 | Stratigraphy of phyllosilicate-bearing strata in the Nili Fossae region.** **a**, OMEGA phyllosilicate detections (green) based on the D2300 parameter<sup>11</sup> overlaid on a Viking digital image mosaic. Targeted CRISM images are outlined; their full identifications are preceded by FRT0000. **b**, CRISM scene FRT00003E12 overlaid with CRISM mineral indicators of olivine (red) and phyllosilicate (blue). The black box indicates the location of **c**. **c**, A portion of HiRISE image PSP\_002176\_2025. Phyllosilicates are found in the bright layered material, overlain by a thin (one or two CRISM pixels) layer of olivine capped by a spectrally neutral unit. **d**, Mineral indicators from CRISM observations FRT000064D9 and FRT00007BC8 overlaid on CTX observation P05\_003986\_20\_11\_XI\_21N285W\_070324. Low-Ca pyroxene is shown in green, phyllosilicate in blue and magenta, and olivine in red. The boundary between Syrtis Major lava and phyllosilicate-bearing material within Nili Fossae trough material is shown by the white arrows.

Phyllosilicates have also been identified by CRISM within sedimentary basins, specifically in fans and deltas within Holden, Eberswalde and Jezero craters, and represent the detection on Mars of hydrated silicates within sediments clearly deposited by water. The mineralogy of phyllosilicates within the sedimentary deposits does not differ from that in nearby potential sediment source regions, whether interior to the basin, as in Holden<sup>22,23</sup>, or from Jezero crater's exterior catchment<sup>24</sup>. The sedimentary phyllosilicates may have been eroded and transported rather than being formed *in situ*, although formation *in situ* cannot be ruled out; it is likely that some alteration of materials within the crater occurred during the various periods of aqueous activity. However, there is a relative paucity of phyllosilicate spectral signatures in the uppermost units of the sedimentary deposits, implying that conditions favouring phyllosilicate formation, deposition or preservation did not persist into the late stages of delta formation in the Noachian.

The CRISM observations and coordinated CTX and HiRISE imaging showed detailed geological relationships and compositional stratigraphy. In the Nili Fossae and Isidis regions we observe regionally extensive phyllosilicate-bearing rocks overlain by both unaltered rocks susceptible to alteration (olivine, pyroxene) and/or a spectrally neutral mesa-forming unit and embayed by Hesperian Syrtis Major

lavas that show no evidence of alteration. Some of the rock units in this region are cut by fractures associated with the Isidis impact basin, which is dated to the late Noachian. Because the olivine-bearing units were emplaced before or concurrently with the event that formed the Isidis basin<sup>17,19</sup>, the lack of any aqueous alteration associated with the olivine lithologies, and the lack of alteration of Hesperian lavas, definitively document Noachian phyllosilicate formation.

The Mawrth Vallis and Nili Fossae regions show that thick, regionally extensive sections of Noachian crust are altered. Throughout the southern highlands, CRISM data document phyllosilicates in the walls, ejecta and central peaks of craters in the southern highlands. This striking association of phyllosilicates with impact craters may be the result of impacts excavating pre-existing deposits or impact-associated processes, leading to the formation of phyllosilicate and hydrated silicate minerals. The possibility cannot be excluded that the alteration occurred as a result of the impact hydrothermal processes<sup>25,26</sup>. However, these systems are usually restricted to the subsurface and/or interior of craters (for example melt sheets). The uniform distribution of phyllosilicate-rich material in crater ejecta, and the presence of phyllosilicates in the walls, floor deposits and within central peaks, argue for the presence of hydrated silicates in the target rocks before impact. With hydrated silicates observed in central peaks of impact



**Figure 4 | Phyllosilicate occurrences in the Noachian-aged southern highlands.** **a**, Map of phyllosilicate mineral indicator<sup>11</sup> (green, weak; orange, strong) derived from CRISM multispectral mapping merged with a MOC wide-angle camera image mosaic. White arrows show distinct regions enriched in phyllosilicate. **b**, High-resolution CRISM observation

FRT000049BB at 13.25° S, 105.25° E, showing a phyllosilicate mineral indicator map (blue, low; green, high). **c**, High-resolution CRISM observation FRT00003E92 at 13.5° S, 119.5° E, showing phyllosilicate (cyan) and pyroxene (red) mineral indicator maps.



craters on the order of 45 km, a significant thickness of the ancient martian crust has experienced some form of alteration. In contrast, despite many observations, no craters in terrains of Hesperian age or in the northern plains where subsurface ice is known to be present show evidence for phyllosilicates.

Identifying the types of phyllosilicate minerals on Mars also constrains their environment of formation. The majority of phyllosilicate spectra are most consistent with smectite clay minerals such as montmorillonite, nontronite and saponite. Smectite clays require significant water reservoirs and moderate to alkaline pH (ref. 3). However, the higher spatial resolution of CRISM has augmented the phyllosilicate mineral diversity beyond this class alone. CRISM has expanded the geographic range of Al-OH-bearing phyllosilicates with a new detection in the Nili Fossae region, where spectra are more consistent with illite or muscovite and kaolinite rather than with montmorillonite. In addition, several regions show absorption features consistent with chlorite. Overall, the weathering products of Fe/Mg mafic minerals are over-represented relative to alkaline minerals such as plagioclase, given the basaltic mineralogy of the martian crust. Kaolinite, which would be expected from an active hydrological system operating in regions with good drainage and high levels of flushing<sup>27</sup>, has been identified in only a few locations, and exposures are less than 200 m × 200 m in spatial scale. The restriction of kaolinite to such rare, small deposits argues against a widespread and vigorous hydrological system in the Noachian or the possibility that the lifetime of such a system was sufficient for the formation of smectite clay but not kaolinite. However, the existence of localities with kaolinite and other diverse phases such as chlorite and hydrated silica suggests that microenvironments with a greater throughput of liquid water or enhanced hydrothermal activity were also a feature of early Mars.

## METHODS SUMMARY

Standard processing approaches were used to convert CRISM data from instrument units to apparent  $I/F$  (the ratio of reflected intensity to incident intensity of sunlight) (ref. 5). We then assume a lambertian surface and divide the data by the cosine of the incidence angle. We assume that surface and atmospheric contributions are multiplicative and that the atmospheric contribution follows an exponential variation with altitude<sup>28</sup>. The data are divided by a scaled, atmospheric transmission spectrum obtained from an observation across Olympus Mons, as used by the OMEGA instrument team<sup>1,15</sup>. This method removes only the effects of gas absorption. However, it is efficient for the analysis of large volumes of data and compares favourably with a more rigorous but time-intensive retrieval of surface Lambert albedos using DISORT-based simulations and regressions for each wavelength band between modelled CRISM radiances and a suite of input surface albedos<sup>29</sup> that addresses both gas absorptions and aerosols. Spectral summary parameters are then derived for each observation, to highlight areas with interesting mineralogy<sup>30</sup>.

Narrow absorptions in the 1.9–2.5- $\mu$ m wavelength region are due to combinations of molecular vibrations in phyllosilicate minerals. Residual calibration artefacts and errors in atmospheric removal cause systematic spikes (Fig. 1a) that are suppressed in spectral ratios in which the spectrum for a region of interest is divided by the spectrum of a nearby region of low spectral contrast. If water is present in a mineral (for example interlayer water in smectite clays), then an absorption centred near 1.9  $\mu$ m is observed as a result of the combination of the H–O–H bending and stretching vibrations<sup>29</sup>. Absorptions near 2.2  $\mu$ m are indicative of Al–OH vibrations, whereas those between 2.28 and 2.35  $\mu$ m are commonly indicative of Fe/Mg–OH vibrations<sup>8–10</sup>. Quantitative abundances can be estimated from computationally intensive, nonlinear spectral deconvolution<sup>31</sup>. Thus, our analyses focus on the detection of phyllosilicate mineralogy, and a qualitative measure of relative abundance is provided by the spectral parameters<sup>30</sup>.

Received 24 April; accepted 8 May 2008.

Published online 16 July 2008.

1. Bibring, J. P. *et al.* Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science* **307**, 1576–1581 (2005).
2. Poulet, F. *et al.* Phyllosilicates on Mars and implications for early martian climate. *Nature* **438**, 623–627 (2005).

3. Velde, B., Righi, D., Meunier, A., Hillier, S. & Inoue, A. in *Origin and Mineralogy of Clays* (ed. Velde, B.) 8–42 (Springer, Berlin, 1995).
4. Bibring, J. P. *et al.* Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. *Science* **312**, 400–404 (2006).
5. Murchie, S. *et al.* Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on Mars Reconnaissance Orbiter (MRO). *J. Geophys. Res.* **112**, doi:10.1029/2006JE002682 (2007).
6. Malin, M. C. *et al.* Context Camera Investigation on board the Mars Reconnaissance Orbiter. *J. Geophys. Res.* **112**, E05S04, doi:10.1029/2006JE002808 (2007).
7. McEwen, A. S. *et al.* Mars Reconnaissance Orbiter's High Resolution Imaging Science Experiment (HiRISE). *J. Geophys. Res.* **112**, E05S02, doi:10.1029/2005JE002605 (2007).
8. Frost, R. L., Klopogge, J. T. & Ding, Z. Near-infrared spectroscopic study of nontronites and ferruginous smectite. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **58**, 1657–1668 (2002).
9. Bishop, J., Madejova, J., Komadel, P. & Froschl, H. The influence of structural Fe, Al and Mg on the infrared OH bands in spectra of dioctahedral smectites. *Clay Miner.* **37**, 607–616 (2002).
10. Clark, R. N., King, T. V. V., Klejwa, M., Swayze, G. A. & Vergo, N. High spectral resolution reflectance spectroscopy of minerals. *J. Geophys. Res. Solid Earth Planets* **95**, 12653–12680 (1990).
11. Calvin, W. M. Could Mars be dark and altered? *Geophys. Res. Lett.* **25**, 1597–1600 (1998).
12. Poulet, F. *et al.* Discovery, mapping and mineralogy of phyllosilicates on Mars by MEX-OMEGA: A reappraisal. *Seventh International Conference on Mars*, abstract 3170 (2007).
13. Loizeau, D. *et al.* Phyllosilicates in the Mawrth Vallis region of Mars. *J. Geophys. Res.* **112**, E08S08, doi:10.1029/2006JE002877 (2007).
14. Michalski, J. R. & Dobrea, E. Z. N. Evidence for a sedimentary origin of clay minerals in the Mawrth Vallis region, Mars. *Geology* **35**, 951–954 (2007).
15. Mustard, J. F. *et al.* Olivine and pyroxene diversity in the crust of Mars. *Science* **307**, 1594–1597 (2005).
16. Hoefen, T. M. *et al.* Discovery of olivine in the Nili Fossae region of Mars. *Science* **302**, 627–630 (2003).
17. Hamilton, V. E. & Christensen, P. R. Evidence for extensive, olivine-rich bedrock on Mars. *Geology* **33**, 433–436 (2005).
18. Mustard, J. F. *et al.* Mineralogy of the Nili Fossae region with OMEGA/Mars Express data: 1. Ancient impact melt in the Isidis Basin and implications for the transition from the Noachian to Hesperian. *J. Geophys. Res.* **112**, E08S03, doi:10.1029/2006JE002834 (2007).
19. Mangold, N. *et al.* Mineralogy of the Nili Fossae region with OMEGA/Mars Express data: 2. Aqueous alteration of the crust. *J. Geophys. Res.* **112**, E08S04, doi:10.1029/2006JE002835 (2007).
20. Bishop, J. L. *et al.* Characterization of phyllosilicates in Libya Montes and the southern Isidis Planitia region using CRISM and HiRISE images. *Seventh International Conference on Mars* (Pasadena, California, abstract 3294 (2007)).
21. Costard, F. *et al.* Detection of hydrated minerals on fluidized ejecta lobes from OMEGA observations: implications in the history of Mars. *37th Lunar and Planetary Science Conference, Houston, Texas* abstract 1288 (2006).
22. Grant, J. A. *et al.* HiRISE imaging of impact megabreccia and sub-meter aqueous strata in Holden crater, Mars. *Geology* **36**, 195–198 (2008).
23. Milliken, R. E. *et al.* Clay minerals in Holden crater as observed by MRO CRISM. *7th International Mars Conference*, abstract 3282 (2007).
24. Ehlmann, B. L. *et al.* Clay minerals in delta deposits and organic preservation potential on Mars. *Nature Geosci.* doi:10.1038/ngeo207 (2008).
25. Newsom, H. E. Hydrothermal alteration of impact melt sheets with implications for Mars. *Icarus* **44**, 207–216 (1980).
26. Rathbun, J. A. & Squyres, S. W. Hydrothermal systems associated with Martian impact craters. *Icarus* **157**, 362–372 (2002).
27. Srodon, J. Nature of mixed-layer clays and mechanisms of their formation and alteration. *Annu. Rev. Earth Planet. Sci.* **27**, 19–53 (1999).
28. Bibring, J. P. *et al.* Results from the ISM Experiment. *Nature* **341**, 591–593 (1989).
29. Wolff, M. J. Constraints on dust aerosols from the Mars Exploration Rovers using MGS overflights and Mini-TES. *J. Geophys. Res.* **111**, E12S17, doi:10.1029/2006JE002786 (2006).
30. Pelkey, S. M. *et al.* CRISM multispectral summary products: Parameterizing mineral diversity on Mars from reflectance. *J. Geophys. Res.* **112**, E08S14, doi:10.1029/2006JE002831 (2007).
31. Poulet, F. & Erard, S. Nonlinear spectral mixing: Quantitative analysis of laboratory mineral mixtures. *J. Geophys. Res.* **109**, E02009, doi:10.1029/2003JE002179 (2004).

**Acknowledgements** We thank the Mars Reconnaissance Orbiter team for building and operating the spacecraft and the numerous people who have contributed to the CRISM investigation. Support from NASA to the CRISM science team is gratefully acknowledged.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.F.M. ([john\\_mustard@brown.edu](mailto:john_mustard@brown.edu)).

## LETTERS

# Generation of Fock states in a superconducting quantum circuit

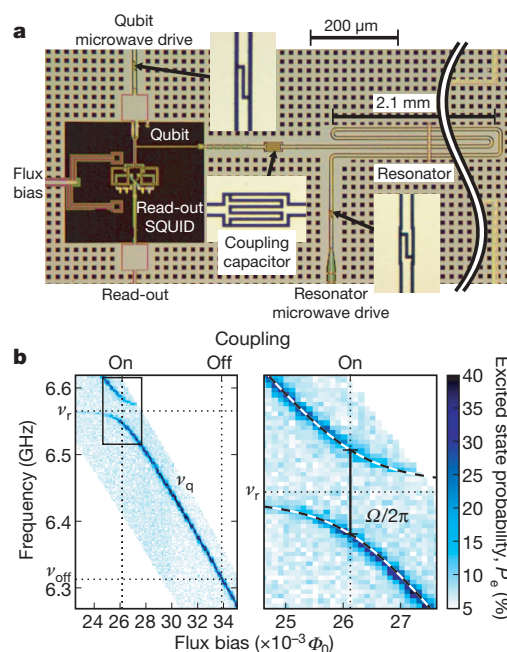
Max Hofheinz<sup>1</sup>, E. M. Weig<sup>1,†</sup>, M. Ansmann<sup>1</sup>, Radosław C. Bialczak<sup>1</sup>, Erik Lucero<sup>1</sup>, M. Neeley<sup>1</sup>, A. D. O'Connell<sup>1</sup>, H. Wang<sup>1</sup>, John M. Martinis<sup>1</sup> & A. N. Cleland<sup>1</sup>

Spin systems and harmonic oscillators comprise two archetypes in quantum mechanics<sup>1</sup>. The spin-1/2 system, with two quantum energy levels, is essentially the most nonlinear system found in nature, whereas the harmonic oscillator represents the most linear, with an infinite number of evenly spaced quantum levels. A significant difference between these systems is that a two-level spin can be prepared in an arbitrary quantum state using classical excitations, whereas classical excitations applied to an oscillator generate a coherent state, nearly indistinguishable from a classical state<sup>2</sup>. Quantum behaviour in an oscillator is most obvious in Fock states, which are states with specific numbers of energy quanta, but such states are hard to create<sup>3–7</sup>. Here we demonstrate the controlled generation of multi-photon Fock states in a solid-state system. We use a superconducting phase qubit<sup>8</sup>, which is a close approximation to a two-level spin system, coupled to a microwave resonator, which acts as a harmonic oscillator, to prepare and analyse pure Fock states with up to six photons. We contrast the Fock states with coherent states generated using classical pulses applied directly to the resonator.

The difficulty of generating quantum number states in a linear resonator has been overcome by interposing a nonlinear quantum system, such as an ion, between a classical radiation source and the resonator. A classical pulse applied to the nonlinear system creates a quantum state that can subsequently be transferred to the resonator. Repeating this process multiple times results in a quantum number state in the resonator. Such a method was used to deterministically generate Fock number states for the mechanical motion of ions in a harmonic ion trap<sup>3</sup>. The analogous deterministic creation of Fock states in electrodynamic resonators has only been demonstrated for states with one or two photons<sup>5,6</sup>, although Fock states with larger photon numbers have been recorded using projective measurements<sup>7,9</sup>. The deterministic creation of pure Fock states in a solid-state system, as described here, represents a significant step forward. Solid-state systems permit highly complex integrated circuitry to employ such bosonic states in, for example, quantum computational architectures. The integration of microwave resonators with solid-state qubits has recently attracted much interest<sup>10–16</sup>, but to date such implementations have only used zero or one photons in the resonator, putting the system in a regime where the bosonic nature of the linear resonator is not apparent.

The method we use here to generate multi-photon Fock states is scalable to arbitrary photon numbers<sup>3,17</sup>, limited only by decoherence times and the speed at which photons can be transferred to the resonator. We generate the Fock states using the qubit as an intermediary between a classical microwave source and the resonator. The Fock states are compared to coherent states generated by driving the resonator directly with a classical radiation pulse. In both cases we

measure the resonator state through the photon number dependence of the qubit–resonator coupling, monitored using the qubit<sup>18</sup>. The complexity of the pulse sequences used to create and analyse the



**Figure 1 | Device description and spectroscopy.** **a**, Photomicrograph of a phase qubit (left) coupled to a co-planar waveguide resonator (right). The resonator has a total length of 8.76 mm. A microwave line capacitively coupled to the qubit is used to inject individual photons into the qubit. A second capacitor between the qubit and resonator couples these two quantum systems, with the resonant interaction controlled by tuning the qubit using a flux bias. The resonator can also be directly excited using a second capacitively coupled microwave line. **b**, Spectroscopy of qubit and resonator. The false-colour images show the excited state probability  $P_e$  of the qubit as a function of driving frequency and flux bias in units of the flux quantum  $\Phi_0 = h/2e$ , where  $e$  is the elementary charge. A dark line is seen when the frequency of the microwave drive matches an eigenfrequency of the qubit–resonator system. An avoided crossing appears when the qubit is tuned through the resonator frequency  $\nu_r$ . A detailed view of the avoided crossing is shown in the right-hand subpanel, superposed with a fit (black-and-white dashed line) to the Jaynes–Cummings model in equation (1). The fitting parameters are the magnitude of the splitting  $\Omega/2\pi = 36.0 \pm 0.6$  MHz (s.e.m. uncertainty; vertical bar) and the resonator frequency  $\nu_r = 6.5666 \pm 0.0005$  GHz (s.e.m. uncertainty; dotted horizontal line). The dotted vertical lines respectively indicate the qubit operating points when the qubit–resonator coupling is on and off.

<sup>1</sup>Department of Physics, University of California, Santa Barbara, California 93106, USA. <sup>†</sup>Present address: Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany.



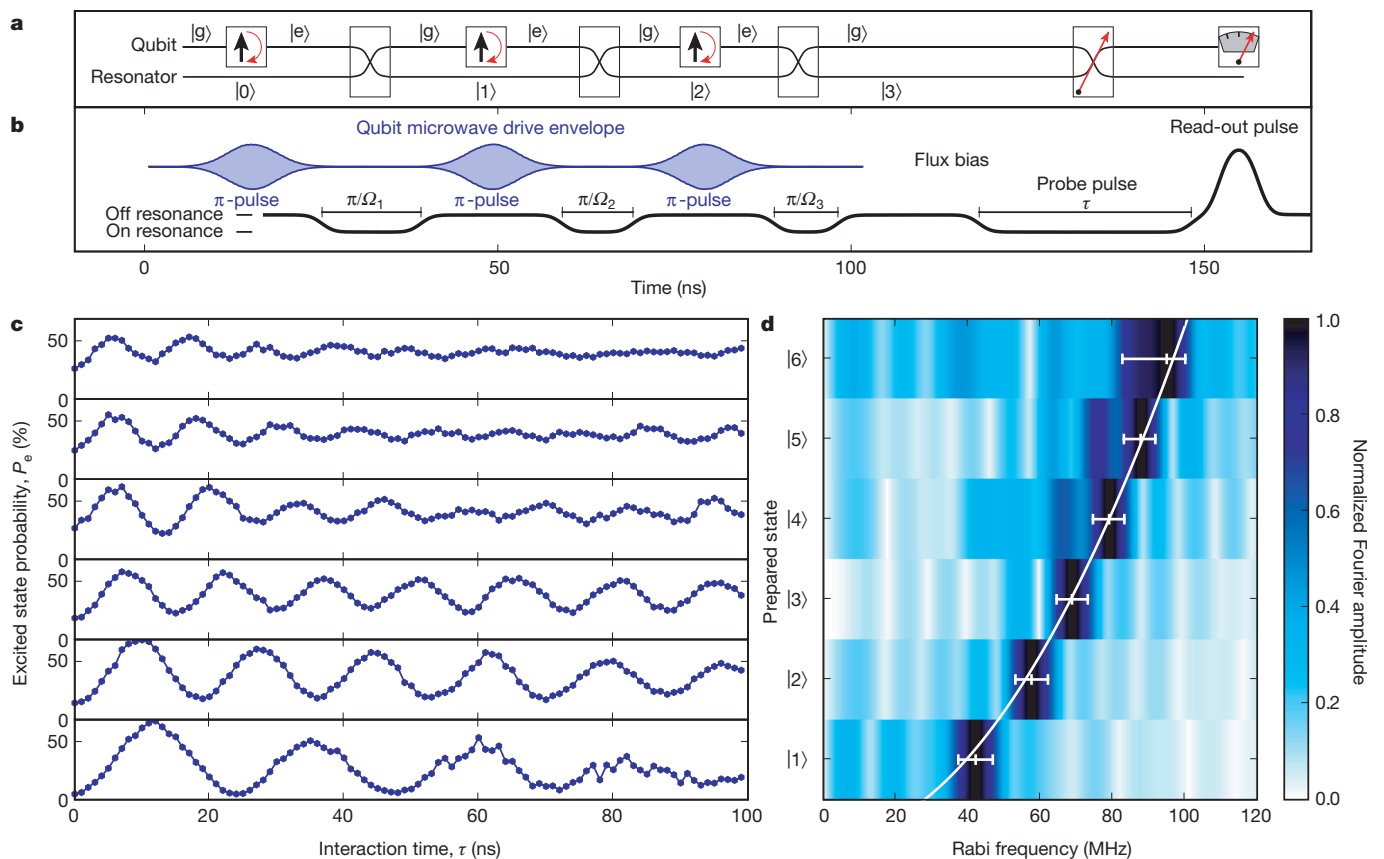
resonator states, and the high fidelity of the resulting measurements, demonstrate a significant advance in the control of superconducting quantum circuits.

Our experimental system (Fig. 1a) is based on the superconducting phase qubit, a device developed for quantum computation<sup>8</sup>. To a good approximation this qubit is represented by a two-level spin system, with ground state  $|g\rangle$  and excited state  $|e\rangle$ . These states are separated in energy by a transition frequency  $\nu_q$  that may be tuned from about 6 to 9 GHz using an external flux bias. With the application of classical microwave pulses, the quantum state of the qubit can be fully controlled<sup>19</sup>. The qubit state is measured by a destructive single-shot measurement, achieved by applying a flux-bias pulse to the qubit. This pulse causes the state  $|e\rangle$  to tunnel to a state that can be easily distinguished from the state  $|g\rangle$  with a flux measurement performed using a read-out d.c. superconducting quantum interference device (SQUID)<sup>20</sup>. The measurement visibility, that is, the difference between the tunnelling probabilities for states  $|e\rangle$  and  $|g\rangle$ , is 80%. Decoherence of the qubit is characterized by measurement of the energy relaxation time  $T_1^q \approx 550$  ns and phase coherence time  $T_2^q \approx 100$  ns.

The qubit is coupled to a high- $Q$ -factor superconducting co-planar waveguide resonator<sup>21</sup>, which serves as a harmonic oscillator, with a resonance frequency  $\nu_r = 6.5666 \pm 0.0005$  GHz. The coupling

is achieved using a capacitor, which sets the coupling strength  $\Omega/2\pi = 36.0 \pm 0.6$  MHz, measured using spectroscopy<sup>22</sup> (Fig. 1b). Achieving strong coupling ( $\Omega \gg 1/T_1$ ) between a phase qubit and a resonator is straightforward<sup>23,24</sup>, as the qubit characteristic impedance of  $\sim 30$  ohms is well matched to the resonator characteristic impedance of  $\sim 50$  ohms. The coupling between the qubit and the resonator can be effectively turned off by biasing the qubit well out of resonance, at a frequency  $\nu_{\text{off}} = 6.314$  GHz, where the coupling is effectively reduced by a factor of  $(\nu_r - \nu_{\text{off}})^2/(\Omega/2\pi)^2 \approx 50$ . Microwaves can also be injected directly into the resonator through a separate microwave feed line. The decoherence times of the resonator were measured to be  $T_1^r \approx 1$   $\mu$ s and  $T_2^r \approx 2$   $\mu$ s  $\approx 2T_1^r$  (E. M. Weig *et al.*, manuscript in preparation). All measurements were performed in a dilution refrigerator operating at 25 mK, which is much less than  $h\nu_r/k_B$  and  $h\nu_q/k_B$  (where  $k_B$  and  $h$  are the Boltzmann and Planck constants, respectively), so thermal noise in this system is negligible.

When the qubit and resonator are tuned off resonance, such that  $|\nu_r - \nu_q| \gg \Omega/2\pi$ , no photons are exchanged between the qubit and resonator. On resonance, for  $|\nu_r - \nu_q| \ll \Omega/2\pi$ , energy can be exchanged between the two systems, and the state of the system can oscillate. The dynamics of energy exchange between the resonator and qubit can be approximated within the rotating-wave



**Figure 2 | Preparation and measurement of Fock states.** **a**, Quantum program and **b**, pulse sequence for the qubit microwave signal and flux bias used to implement it. An excitation is created in the qubit with a resonant microwave pulse and then transferred to the resonator by tuning the qubit into resonance for half an oscillation period. This sequence is repeated until the desired photon number  $n$  is reached;  $n = 3$  in the example depicted here. The length of the tuning pulse decreases as  $1/\sqrt{n}$ . To analyse the resonator state the qubit is tuned into resonance for a variable interaction time  $\tau$  and the qubit state is finally read out by applying a high-flux-bias pulse that makes the excited state tunnel into a state which can be easily distinguished from the ground state. **c**, Plot of the probability of the excited qubit state  $P_e$  versus interaction time  $\tau$  for Fock states with  $n = 1, \dots, 6$ . The time traces

show sinusoidal oscillations with a period that shortens with increasing photon number  $n$ . **d**, The false-colour image shows the Fourier amplitudes of the traces in **c**, obtained with a 100-ns rectangular window function after subtracting the average value. Each Fourier transform displays a clear peak at the  $n$ -photon oscillation frequency  $\Omega_n$ , indicating the high purity of the Fock states. The peak maxima are marked, with the error bars indicating their  $-3$ -dB points. The white curve is the expected  $\sqrt{n}$  scaling, adjusted to fit the data using the coupling strength  $\Omega/2\pi = 40 \pm 1$  MHz (s.e.m. uncertainty), which is slightly higher than that determined from spectroscopy. In addition, the actual photon number scaling of the oscillation frequency is slightly slower than  $\sqrt{n}$ . Both deviations can be attributed to detuning of the qubit with respect to the resonator, as discussed in the text.

approximation by the Jaynes–Cummings model Hamiltonian<sup>25</sup>

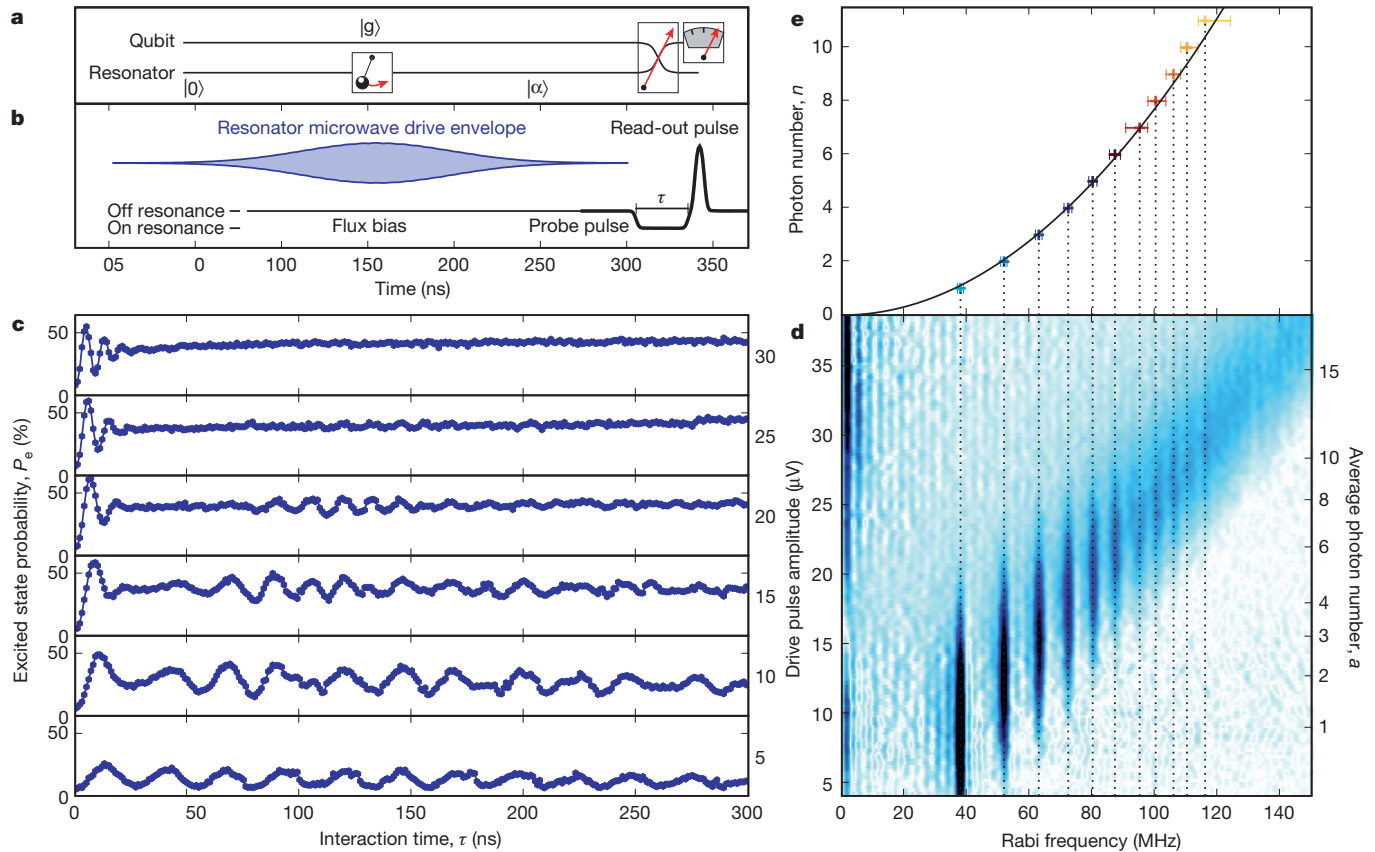
$$H_{\text{int}} = \frac{\hbar\Omega}{2} (a\sigma_+ + a^\dagger\sigma_-) \quad (1)$$

where  $a^\dagger$  and  $a$  are respectively the photon creation and annihilation operators for the resonator,  $\sigma_+$  and  $\sigma_-$  are respectively the qubit raising and lowering operators and  $\hbar = h/2\pi$ . If the system is prepared in the state  $|g\rangle|n\rangle$  (qubit in the ground state,  $n$  photons in the resonator), the system will oscillate between this state and the state  $|e\rangle|n-1\rangle$  at an angular frequency  $\Omega_n = \sqrt{n}\Omega$ . This  $\sqrt{n}$  dependence of the oscillation frequency is the cavity quantum electrodynamic equivalent of stimulated emission: a photon is transferred between resonator and qubit more rapidly when more photons are present in the resonator. This increase in oscillation frequency is the key to our measurement of the resonator state<sup>18</sup>.

In order to prepare the resonator in a Fock number state, we begin with the qubit detuned from the resonator and wait a time much longer than  $T_1^r$  or  $T_1^q$ , allowing both qubit and resonator to relax to their respective ground states  $|g\rangle$  and  $|0\rangle$ . As shown in Fig. 2a, b, we then apply a gaussian microwave pulse to the qubit at  $\nu_q = \nu_{\text{off}}$  with

an amplitude and duration that are calibrated to yield the state  $|e\rangle$ . We obtain  $\sim 98\%$  fidelity for this operation when implemented with properly shaped pulses<sup>26</sup>. The qubit and resonator are then tuned into resonance for a time  $\pi/\Omega_1 = \pi/\Omega$  so that the excitation in the qubit is transferred to the resonator. The time and amplitude of the tuning pulse are adjusted to yield the best state transfer, determined by maximizing the probability of finding the qubit in the state  $|g\rangle$  directly after the pulse. A second microwave pulse is then applied to the qubit to re-prepare it in the state  $|e\rangle$ . The qubit and resonator are brought back into resonance, but for a reduced time  $\pi/\Omega_2 = \pi/\sqrt{2}\Omega$ . After this procedure has been repeated  $n$  times, with an appropriate reduction in the transfer time for each successive photon, we obtain a final state  $|g\rangle|n\rangle$  corresponding to an  $n$ -photon Fock state in the resonator.

To analyse the resonator state, we tune the qubit and resonator into resonance for an adjustable interaction time  $\tau$  and then read out the qubit state. The probability  $P_e(\tau)$  of finding the qubit in the state  $|e\rangle$  is obtained by averaging 3,000 pulse sequences for each interaction time  $\tau$ . The probability is expected to oscillate in the absence of decoherence according to<sup>2</sup>



**Figure 3 | Preparation and measurement of coherent states.** **a**, Quantum program and **b**, pulse sequence of the resonator microwave drive and the qubit bias used to implement it. A gaussian pulse with 100-ns full-width at half-maximum and varying amplitude is directly applied to the resonator and creates a coherent state  $|\alpha\rangle$ . The qubit, in its ground state, is then brought into resonance for a variable interaction time  $\tau$  and measured, exactly as for the Fock state measurements. **c**, Plot of the excited state probability  $P_e$  versus interaction time  $\tau$  for six different microwave amplitudes. The time traces are aperiodic because of the irrational ratios in the oscillation times for the different photon number states  $|n\rangle$  comprising the coherent state. **d**, Fourier transform of the data in **c**, obtained with a 300-

ns rectangular window function after subtracting the average value. Darker colours indicate higher amplitudes. The data have been smoothed in the drive pulse direction with a  $\sigma = 0.2$ - $\mu\text{V}$  gaussian low-pass filter. The Fourier spectrum reveals a sharp peak at each number state frequency for  $n = 1, \dots, 11$ . **e**, Frequencies of the  $n$ -photon Fourier peaks compared to the expected  $\sqrt{n}$  scaling, adjusted to fit the data using the coupling strength  $\Omega/2\pi = 36.0 \pm 0.3$  MHz (s.e.m. uncertainty), which matches that determined from spectroscopy. As in Fig. 2d, the actual scaling is slightly slower than  $\sqrt{n}$ . The error bars indicate the  $-3$ -dB points of the Fourier transform peaks.



$$P_e(\tau) = \sum_{n=1}^{\infty} P_n \frac{1 - \cos(\Omega_n \tau)}{2} \quad (2)$$

where  $P_n$  is the probability of initially having  $n$  photons in the resonator. For a pure Fock state  $|n\rangle$ ,  $P_e$  oscillates at the frequency  $\Omega_n/2\pi$ . When several different  $|n\rangle$  states are occupied, the time dependence of  $P_e(\tau)$  becomes more complex, owing to the irrational ratios of the oscillation frequencies  $\Omega_n$ . Note that although  $P_0$  does not enter equation (2) directly, it is given by the time average

$$\bar{P}_e = \sum_{n=1}^{\infty} \frac{P_n}{2} = \frac{1 - P_0}{2} \quad (3)$$

The experimental time dependence of  $P_e(\tau)$  is displayed in Fig. 2c for Fock states with  $n = 1, \dots, 6$ . The time traces are approximately sinusoidal, indicating from equation (2) that for each initial state, one photon number dominates in the resonator state. The oscillations have large amplitudes for  $n = 1, 2, 3$ , and gradually decrease for  $n = 4, 5, 6$ . Both the amplitude and the decay time decrease with increasing photon number  $n$  because the lifetime of an  $n$ -photon Fock state decreases as  $T_1^n/n$  (ref. 27) and the time needed to create such a state increases as  $n$ . For  $n = 6$ , the lifetime of the Fock state and the length of the preparation sequence are comparable.

The period of the oscillations clearly decreases with  $n$ . The period of the state  $|4\rangle$ , for example, is approximately half the period of the state  $|1\rangle$ , as expected from the  $\sqrt{n}$  scaling of the oscillation frequency. A more quantitative analysis of this dependence is shown in Fig. 2d, where the Fourier transforms of the time traces are plotted: each displays a clear peak at a single frequency, which scales approximately as  $\sqrt{n} \Omega/2\pi$ . The actual frequency dependence is slightly slower than  $\sqrt{n}$ , and the coupling strength  $\Omega/2\pi = 40$  MHz is slightly larger than the one obtained from the splitting in Fig. 1. Both deviations can be explained by our having used an 'on' operating point slightly detuned from the minimal splitting in Fig. 1b, yielding slightly higher oscillation frequencies. The oscillation frequencies for higher photon number states, however, are less affected by detuning because they are more strongly coupled to the qubit. Indeed, for this experiment we choose the on-point such as to maximize state transfer between the qubit and the resonator. This on-point is slightly off resonance owing to imperfections in the tuning pulses.

We next highlight the non-classical features of the Fock states by comparing them with coherent states, the quantum equivalent of classical oscillations, that are created when a harmonic oscillator is driven directly with a classical signal. To generate such states we drive the resonator with a gaussian-shaped resonant microwave pulse with a full-width at half-maximum of 100 ns (Fig. 3a, b) and a range of amplitudes. The qubit is not involved in this state preparation and stays in the ground state. The read-out of the resonator state is performed using the qubit, exactly as for the Fock state analysis.

In a coherent state the amplitude and phase of the oscillation are well defined, but the number of photons is not. A coherent state is a superposition of different Fock states for which the occupation probability  $P_n$  of an  $n$ -photon Fock state follows a Poisson distribution that depends on the average photon number  $a$ :

$$P_n(a) = \frac{a^n e^{-a}}{n!} \quad (4)$$

As a result, the time dependence of  $P_e(\tau)$  (Fig. 3c) is strikingly different from that observed for the Fock states. At low drive amplitude,  $P_e(\tau)$  is periodic but has low visibility because for  $a \ll 1$  all  $P_n$  for  $n > 1$  are vanishingly small and  $P_1$  itself is small. At higher drive amplitudes, the time traces display a strong initial ringing with fast collapse, followed by a revival—a characteristic feature of a coherent state coupled to a two-level system<sup>28,29</sup>. During the revival, the time dependence is irregular because the coherent state is composed of different Fock states that oscillate with irrational frequency ratios.

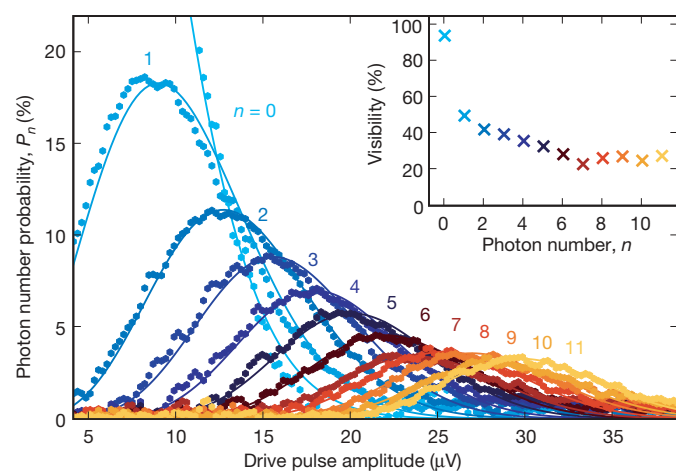
The decomposition of the coherent states into Fock states becomes very clear in the Fourier transforms of the time traces (Fig. 3d). The

oscillation frequencies for different photon number states appear as sharp vertical lines, indicating the underlying quantum nature of the coherent states. With increasing pulse amplitude, the lines corresponding to higher photon numbers become more pronounced, and at any given pulse amplitude there are several sequential photon number states with significant occupation probabilities. In Fig. 3e, the oscillation frequencies corresponding to the maxima of these lines are plotted versus the corresponding photon number. The dependence on photon number matches that observed previously in the analysis of the Fock states.

These data also show good quantitative agreement with the expected Poisson distributions. In Fig. 4 we plot the photon number state probabilities  $P_n$  obtained from the Fourier amplitudes along the dashed vertical lines in Fig. 3d. Their dependence on drive amplitude agrees very well with the Poisson distributions, which are plotted as solid lines. The Poisson distribution for each photon number  $n$  has been scaled by a visibility (Fig. 4 inset). The visibility for  $n = 0$  is close to 100%. The visibility for higher photon number states is lower because the Fourier amplitude is reduced by decoherence during the interaction time of 300 ns. We find that shorter interaction times yield much higher visibilities, but at the cost of lower frequency resolution in Fig. 3d, e.

We note that, unlike that for a pure Fock state, the photon number distribution for a coherent state does not reveal the entire quantum description of the resonator state, because a number state analysis cannot by itself distinguish a statistical mixture from a pure coherent state. A full quantum analysis would involve a complete tomographic measurement, yielding, for example, the Wigner function<sup>30</sup>. Given the high fidelity of our Fock state measurements and the excellent agreement with the coherent state analysis, we believe that such an experiment should be possible in the near future.

In conclusion, we have created multi-photon Fock states for the first time in a solid-state system. The highest photon number we have achieved to date,  $n = 6$ , is limited only by the coherence times of the qubit and the resonator. In our experiment, the Fock states are created on demand in a completely deterministic fashion. This presents the possibility of using complex bosonic states in solid-state-based quantum algorithms, which until now have only involved spin-1/2-like (fermionic) states.



**Figure 4 | Population analysis of coherent states.** The points represent the amplitudes of the Fourier transforms in Fig. 3d, obtained along the dashed vertical lines, corrected for the measurement visibility of 80% and transformed into photon number probabilities using equations (2) and (3). The solid lines are the photon number probabilities predicted by the Poisson distribution in equation (4). To fit the measured distribution, the solid lines have been scaled by an individual visibility factor for each photon number (inset). The visibility for  $n = 0$  is close to 100%; the visibilities for  $n > 0$  are lower because decoherence during the measurement lowers the corresponding Fourier amplitudes.

Received 19 March; accepted 23 May 2008.

1. Cohen-Tannoudji, C., Diu, B. & Laloë, F. *Quantum Mechanics* Vol. 1, Ch. 4-5 (Wiley, New York, 2006).
2. Haroche, S. & Raimond, J.-M. *Exploring the Quantum — Atoms, Cavities and Photons* Ch. 3 (Oxford Univ. Press, Oxford, UK, 2006).
3. Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
4. Cirac, J. I., Blatt, R., Parkins, A. S. & Zoller, P. Preparation of Fock states by observation of quantum jumps in an ion trap. *Phys. Rev. Lett.* **70**, 762–765 (1993).
5. Varcoe, B. T. H., Brattke, S., Weidinger, M. & Walther, H. Preparing pure photon number states of the radiation field. *Nature* **403**, 743–746 (2000).
6. Bertet, P. *et al.* Generating and probing a two-photon Fock state with a single atom in a cavity. *Phys. Rev. Lett.* **88**, 143601 (2002).
7. Waks, E., Dimanti, E. & Yamamoto, Y. Generation of photon number states. *N. J. Phys.* **8**, 4 (2006).
8. Devoret, M. & Martinis, J. M. Implementing qubits with superconducting integrated circuits. *Quantum Inf. Process.* **3**, 163–203 (2004).
9. Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
10. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
11. Johansson, J. *et al.* Vacuum Rabi oscillations in a macroscopic superconducting qubit LC oscillator system. *Phys. Rev. Lett.* **96**, 127006 (2006).
12. Houck, A. A. *et al.* Generating single microwave photons in a circuit. *Nature* **449**, 328–331 (2007).
13. Sillanpää, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449**, 438–442 (2007).
14. Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
15. Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
16. Astafiev, O. *et al.* Single artificial-atom lasing. *Nature* **449**, 588–590 (2007).
17. Liu, Y.-X., Wei, L. F. & Nori, F. Generation of nonclassical photon states using a superconducting qubit in a microcavity. *Europhys. Lett.* **67**, 941–947 (2004).
18. Brune, M. *et al.* Quantum Rabi oscillation: A direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
19. Steffen, M. *et al.* State tomography of capacitively shunted phase qubits with high fidelity. *Phys. Rev. Lett.* **97**, 050502 (2006).
20. Neeley, M. *et al.* Transformed dissipation in superconducting quantum circuits. *Phys. Rev. B* **77**, 180508(R) (2008).
21. O'Connell, A. D. *et al.* Microwave dielectric loss at single photon energies and millikelvin temperatures. *Appl. Phys. Lett.* **92**, 112903 (2008).
22. Neeley, M. *et al.* Process tomography of quantum memory in a Josephson phase qubit coupled to a two-level state. *Nature Phys.* advance online publication doi:10.1038/nphys972 (27 April 2008).
23. Devoret, M. H., Esteve, D., Martinis, J. M. & Urbina, C. Effect of an adjustable admittance on the macroscopic energy levels of a current biased Josephson junction. *Phys. Scr.* **T25**, 118–121 (1989).
24. Devoret, M. H. *et al.* in *Quantum Tunnelling in Condensed Media* (eds Kagan, Y. & Leggett, A. J.) Ch. 6 337–338 (Elsevier, Amsterdam, 1992).
25. Jaynes, E. & Cummings, F. Comparison of quantum and semiclassical radiation theories with application to the beam maser. *Proc. IEEE* **51**, 89–109 (1963).
26. Lucero, E. *et al.* High-fidelity gates in a Josephson qubit. *Phys. Rev. Lett.* (in the press); preprint at (<http://arxiv.org/abs/0802.0903>) (2008).
27. Lu, N. Effects of dissipation on photon statistics and the lifetime of a pure number state. *Phys. Rev. A* **40**, 1707–1708 (1989).
28. Faist, A., Geneux, E., Meystre, P. & Quattropani, P. Coherent radiation in interaction with two-level system. *Helv. Phys. Acta* **45**, 956–959 (1972).
29. Eberly, J. H., Narozhny, N. B. & Sanchez-Mondragon, J. J. Periodic spontaneous collapse and revival in a simple quantum model. *Phys. Rev. Lett.* **44**, 1323–1326 (1980).
30. Leibfried, D. *et al.* Experimental determination of the motional quantum state of a trapped atom. *Phys. Rev. Lett.* **77**, 4281–4285 (1996).

**Acknowledgements** We thank M. Geller for theoretical input. Devices were made at the UCSB Nanofabrication Facility, a part of the NSF-funded National Nanotechnology Infrastructure Network. This work was supported by IARDA under grant W911NF-04-1-0204 and by the NSF under grants CCF-0507227 and DMR-0605818.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.N.C. ([cleland@physics.ucsb.edu](mailto:cleland@physics.ucsb.edu)).



# Climbing the Jaynes–Cummings ladder and observing its $\sqrt{n}$ nonlinearity in a cavity QED system

J. M. Fink<sup>1</sup>, M. Göppl<sup>1</sup>, M. Baur<sup>1</sup>, R. Bianchetti<sup>1</sup>, P. J. Leek<sup>1</sup>, A. Blais<sup>2</sup> & A. Wallraff<sup>1</sup>

The field of cavity quantum electrodynamics (QED), traditionally studied in atomic systems<sup>1–3</sup>, has gained new momentum by recent reports of quantum optical experiments with solid-state semiconducting<sup>4–8</sup> and superconducting<sup>9–11</sup> systems. In cavity QED, the observation of the vacuum Rabi mode splitting is used to investigate the nature of matter–light interaction at a quantum-mechanical level. However, this effect can, at least in principle, be explained classically as the normal mode splitting of two coupled linear oscillators<sup>12</sup>. It has been suggested that an observation of the scaling of the resonant atom–photon coupling strength in the Jaynes–Cummings energy ladder<sup>13</sup> with the square root of photon number  $n$  is sufficient to prove that the system is quantum mechanical in nature<sup>14</sup>. Here we report a direct spectroscopic observation of this characteristic quantum nonlinearity. Measuring the photonic degree of freedom of the coupled system, our measurements provide unambiguous spectroscopic evidence for the quantum nature of the resonant atom–field interaction in cavity QED. We explore atom–photon superposition states involving up to two photons, using a spectroscopic pump and probe technique. The experiments have been performed in a circuit QED set-up<sup>15</sup>, in which very strong coupling is realized by the large dipole coupling strength and the long coherence time of a superconducting qubit embedded in a high-quality on-chip microwave cavity. Circuit QED systems also provide a natural quantum interface between flying qubits (photons) and stationary qubits for applications in quantum information processing and communication<sup>16</sup>.

The dynamics of a two-level system coupled to a single mode of an electromagnetic field is described by the Jaynes–Cummings hamiltonian:

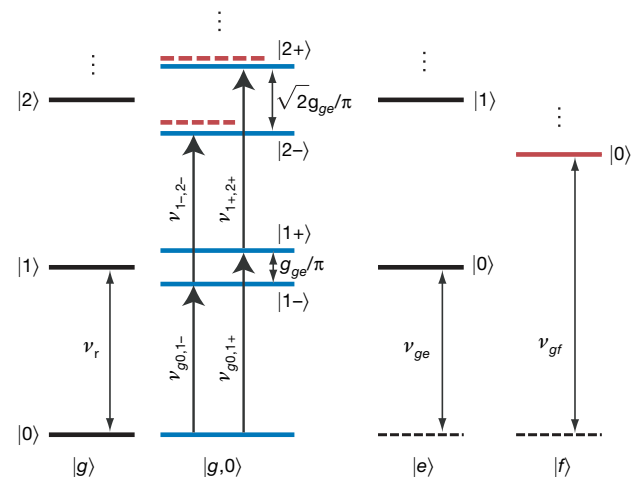
$$\hat{H}_0 = \hbar\omega_{ge}\hat{\sigma}_{ee} + \hbar\omega_r a^\dagger a + \hbar g_{ge}(\hat{\sigma}_{ge}^\dagger a + a^\dagger \hat{\sigma}_{ge}) \quad (1)$$

Here  $\omega_{ge}$  is the transition frequency between the ground  $|g\rangle$  and excited state  $|e\rangle$  of the two-level system,  $\omega_r$  is the frequency of the resonator field and  $g_{ge}$  is the coupling strength between the two.  $a^\dagger$  and  $a$  are the raising and lowering operators acting on the photon number states  $|n\rangle$  of the field, and  $\hat{\sigma}_{ij} = |i\rangle\langle j|$  are the corresponding operators acting on the qubit states. When the coherent coupling rate  $g_{ge}$  is larger than the rate  $\kappa$  at which photons are lost from the field and larger than the rate  $\gamma$  at which the two-level system loses its coherence, the strong-coupling limit is realized. On resonance ( $\omega_{ge} = \omega_r$ ) and in the presence of  $n$  excitations, the new eigenstates of the coupled system are the symmetric  $(|g,n\rangle + |e,n-1\rangle)/\sqrt{2} \equiv |n+\rangle$  and anti-symmetric  $(|g,n\rangle - |e,n-1\rangle)/\sqrt{2} \equiv |n-\rangle$  qubit–photon superposition states (Fig. 1). For  $n = 1$ , these states are equivalently observed spectroscopically as a vacuum Rabi mode splitting<sup>4–9,17,18</sup> or in time resolved measurements as vacuum Rabi oscillations<sup>11,19–21</sup> at frequency  $2g_{ge}$ . The Jaynes–Cummings model predicts a characteristic nonlinear scaling of this frequency as  $\sqrt{n}2g_{ge}$  with the number of

excitations  $n$  in the system (Fig. 1). This quantum effect is in stark contrast to the normal mode splitting of two classical coupled linear oscillators, which is independent of the oscillator amplitude.

Since the first measurements of the vacuum Rabi mode splitting with, on average, a single intracavity atom<sup>17</sup>, it remains a major goal to clearly observe this characteristic  $\sqrt{n}$  nonlinearity spectroscopically to prove the quantum nature of the interaction between the two-level system and the radiation field<sup>12,14,22</sup>. In time domain measurements of vacuum Rabi oscillations, evidence for this  $\sqrt{n}$  scaling has been found with circular Rydberg atoms<sup>19</sup> and superconducting flux qubits<sup>11</sup> interacting with weak coherent fields. Related experiments have been performed with one- and two-photon Fock states<sup>20,21</sup>. We now observe this nonlinearity directly using a scheme similar to the one suggested in ref. 22 by pumping the system selectively into the first doublet  $|1\pm\rangle$  and probing transitions to the second doublet  $|2\pm\rangle$ . This technique realizes efficient excitation into higher doublets at small intracavity photon numbers, avoiding unwanted a.c. Stark shifts that occur in high-drive and elevated-temperature experiments.

In a different regime, when the qubit is detuned by an amount  $|\Delta| = |\omega_{ge} - \omega_r| \gg g_{ge}$  from the cavity, photon number states and their distribution have recently been observed using dispersive quantum non-demolition measurements in both circuit QED<sup>23</sup> and Rydberg atom experiments<sup>24</sup>.

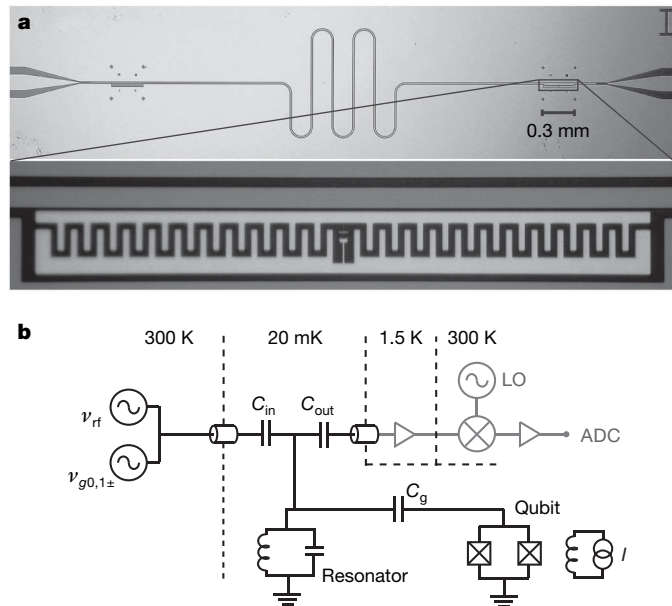


**Figure 1 | Level diagram of a resonant ( $\nu_r = \nu_{ge}$ ) cavity QED system.** The uncoupled qubit states  $|g\rangle$ ,  $|e\rangle$  and  $|f\rangle$  from left to right and the photon states  $|0\rangle$ ,  $|1\rangle$  and  $|2\rangle$  from bottom to top are shown. Vertical dots signify the continuation of the Jaynes–Cummings ladder to larger  $|n\rangle$ . The dipole coupled dressed states are shown in blue and a shift due to the  $|f, 0\rangle$  level (solid red line) is indicated with dashed red lines. Dashed black lines indicate the ground state energy in this diagram. Pump ( $\nu_{g0,1-}$ ,  $\nu_{g0,1+}$ ) and probe ( $\nu_{1-,2-}$ ,  $\nu_{1+,2+}$ ) transition frequencies are indicated accordingly. See text for details.

<sup>1</sup>Department of Physics, ETH Zürich, CH-8093 Zürich, Switzerland. <sup>2</sup>Département de Physique, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada.

In our experiments, which are in the resonant regime, a superconducting qubit playing the role of an artificial atom is strongly coupled to photons contained in a coplanar waveguide resonator in an architecture known as circuit QED<sup>9,15</sup>. We use a transmon<sup>25,26</sup>, which is a charge-insensitive superconducting qubit design derived from the Cooper-pair box<sup>27</sup>, as the artificial atom. Its transition frequency is given by  $\omega_{ge}/2\pi \approx \sqrt{8E_C E_J(\Phi)}$ , with the single electron charging energy  $E_C = 0.4$  GHz, the flux controlled Josephson energy  $E_J(\Phi) = E_{J,\max} |\cos(\pi\Phi/\Phi_0)|$  and  $E_{J,\max} = 53.5$  GHz, as determined in spectroscopic measurements (here  $\Phi_0$  is the magnetic flux quantum). The cavity is realized as a coplanar resonator with bare resonance frequency  $\nu_r = 6.94$  GHz and decay rate  $\kappa/2\pi = 0.9$  MHz. Optical images of the sample are shown in Fig. 2a. The large dimension of the qubit in the quasi-one-dimensional resonator layout provides a very large dipole coupling strength  $g_{ge}$ . A simplified electrical circuit diagram of the set-up is shown in Fig. 2b.

The system is prepared in its ground state  $|g, 0\rangle$  by cooling it to temperatures below 20 mK in a dilution refrigerator. We then probe the energies of the lowest doublet  $|1\pm\rangle$ , measuring the cavity transmission spectrum  $T$  and varying the detuning between the qubit transition frequency  $\nu_{ge}$  and the cavity frequency  $\nu_r$  by applying a magnetic flux  $\Phi$  (Fig. 3a). The measurement is performed with a weak probe of power  $P \approx -137$  dBm applied to the input port of the resonator, populating it with a mean photon number of  $\bar{n} \approx 1.6$  on resonance when the qubit is maximally detuned from the resonator.  $P$  is calibrated in a dispersive a.c. Stark shift measurement<sup>28</sup>. At half integers of  $\Phi_0$ , the qubit energy level separation  $\nu_{ge}$  approaches

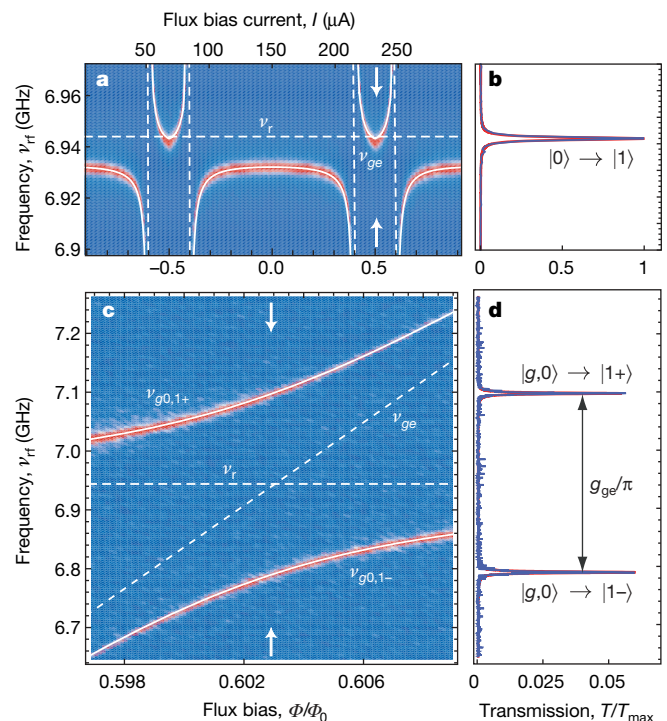


**Figure 2 | Sample and experimental set-up.** **a**, Top, optical image of the superconducting coplanar waveguide resonator with the transmon type superconducting qubit embedded at the position shown boxed. Bottom, magnified view of boxed area, showing the qubit with dimensions  $300 \times 30 \mu\text{m}^2$  close to the centre conductor. **b**, Simplified circuit diagram of the experimental set-up, similar to the one used in ref. 9. The qubit at temperature 20 mK is capacitively coupled to the resonator through  $C_g$ , and the resonator, represented by a parallel LC circuit, is coupled to input and output transmission lines via the capacitors  $C_{\text{in}}$  and  $C_{\text{out}}$ . Flux control is realized with a current biased ( $I$ ) coil close to the qubit. Microwave signal generators for applying pump  $\nu_{g0,1\pm}$  and measurement  $\nu_{\text{rf}}$  tones are indicated. Using ultralow-noise amplifiers at 1.5 K and a mixer at 300 K, the transmitted microwave signal is down-converted with a local oscillator (LO) and digitized with an analog-to-digital converter (ADC).

zero. At this point, the bare resonator spectrum peaked at the frequency  $\nu_r$  is observed (Fig. 3b). We use the measured maximum transmission amplitude to normalize the amplitudes in all subsequent measurements. At all other detunings  $|\Delta| \gg g_{ge}$  the qubit dispersively shifts<sup>25</sup> the cavity frequency  $\nu_r$  by  $\chi \approx -g_{ge}^2 E_C / (\Delta(\Delta - E_C))$ .

Measuring cavity transmission  $T$  as a function of flux bias  $\Phi$  in the anti-crossing region yields transmission maxima at frequencies corresponding to transitions to the first doublet  $|1\pm\rangle$  in the Jaynes–Cummings ladder (Fig. 3c). On resonance ( $\Delta = 0$ ), we extract a coupling strength of  $g_{ge}/2\pi = 154$  MHz (Fig. 3d), where the linewidth of the individual vacuum Rabi split lines is given by  $\delta\nu_0 \approx 2.6$  MHz. This corresponds to a transmission peak separation  $g_{ge}/\pi$  of over 100 linewidths  $\delta\nu_0$ , clearly demonstrating that strong coupling is realized<sup>9,29</sup>. Solid white lines in Fig. 3a, c (and Fig. 4a, c) are numerically calculated dressed state frequencies with the qubit and resonator parameters as stated above, and are in excellent agreement with the data. For the calculation, the qubit hamiltonian is solved exactly in the charge basis. The qubit states  $|g\rangle$  and  $|e\rangle$  and the flux dependent coupling constant  $g_{ge}$  are then incorporated in the Jaynes–Cummings hamiltonian, equation (1). Its numeric diagonalization yields the dressed states of the coupled system without any fit parameters.

In our pump and probe scheme, we first determine spectroscopically the exact energies of the first doublet  $|1\pm\rangle$  at a given flux  $\Phi$ . We then apply a pump tone at the fixed frequency  $\nu_{g0,1-}$  or  $\nu_{g0,1+}$  to populate the respective first doublet state  $|1\pm\rangle$ . A probe tone of the same power is then scanned over the frequency range of the splitting. This procedure is repeated for different flux controlled detunings. The transmission at the pump and probe frequencies is spectrally resolved in a heterodyne detection scheme.



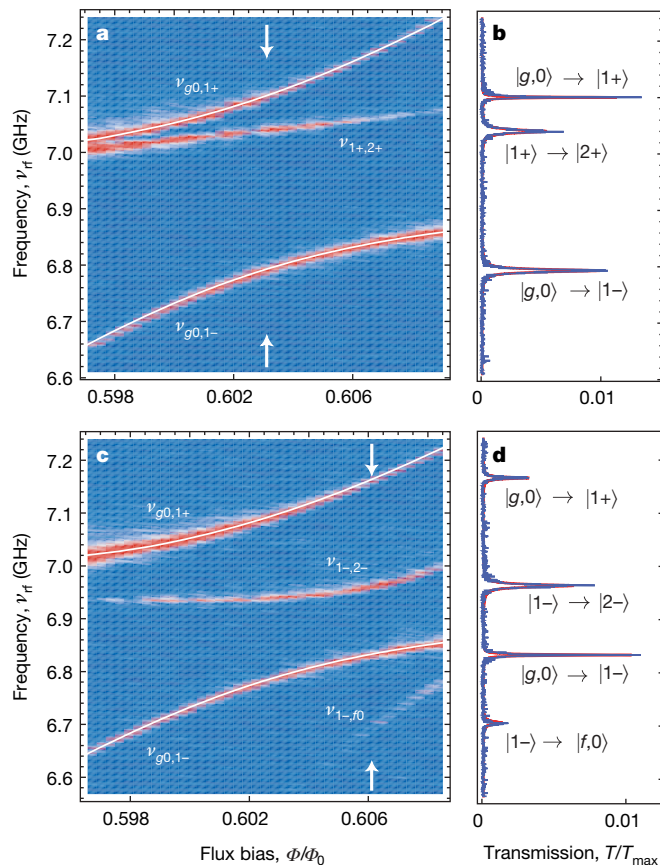
**Figure 3 | Vacuum Rabi mode splitting with a single photon.** **a**, Measured resonator transmission spectra versus normalized external flux bias,  $\Phi/\Phi_0$  (bottom axis) and corresponding bias current  $I$  applied to a superconducting coil (top axis). Transmission  $T$  is colour coded: blue, low; red, high. The solid white line shows dressed state energies as obtained numerically, and the dashed lines indicate the bare resonator frequency  $\nu_r$  as well as the qubit transition frequency  $\nu_{ge}$ . **b**, Normalized resonator transmission  $T/T_{\text{max}}$  at  $\Phi/\Phi_0 = 1/2$ , as indicated with arrows in **a**, with a lorentzian line fit in red. **c**, Resonator transmission  $T$  versus  $\Phi/\Phi_0$  close to degeneracy. **d**, Vacuum Rabi mode splitting at degeneracy, with lorentzian line fit in red.

Populating the symmetric state  $|1+\rangle$ , we observe an additional transmission peak at a probe tone frequency that varies with flux (Fig. 4a). This peak corresponds to the transition between the symmetric doublet states  $|1+\rangle$  and  $|2+\rangle$  at frequency  $\nu_{1+,2+}$ . Similarly, in Fig. 4c, where the antisymmetric state  $|1-\rangle$  is populated, we measure a transmission peak that corresponds to the transition between the two antisymmetric doublet states  $|1-\rangle$  and  $|2-\rangle$  at frequency  $\nu_{1-,2-}$ . The transmission spectra displayed in Fig. 4b, d recorded at the values of flux indicated by arrows in Fig. 4a, c show that the distinct transitions between the different doublets are very well resolved, with separations of tens of linewidths. Transitions between symmetric and antisymmetric doublet states are not observed in this experiment, because the flux-dependent transition matrix elements squared are on average smaller by factors of 10 and 100 for transitions  $|1+\rangle \rightarrow |2-\rangle$  and  $|1-\rangle \rightarrow |2+\rangle$ , respectively, than the corresponding matrix elements between states of the same symmetry.

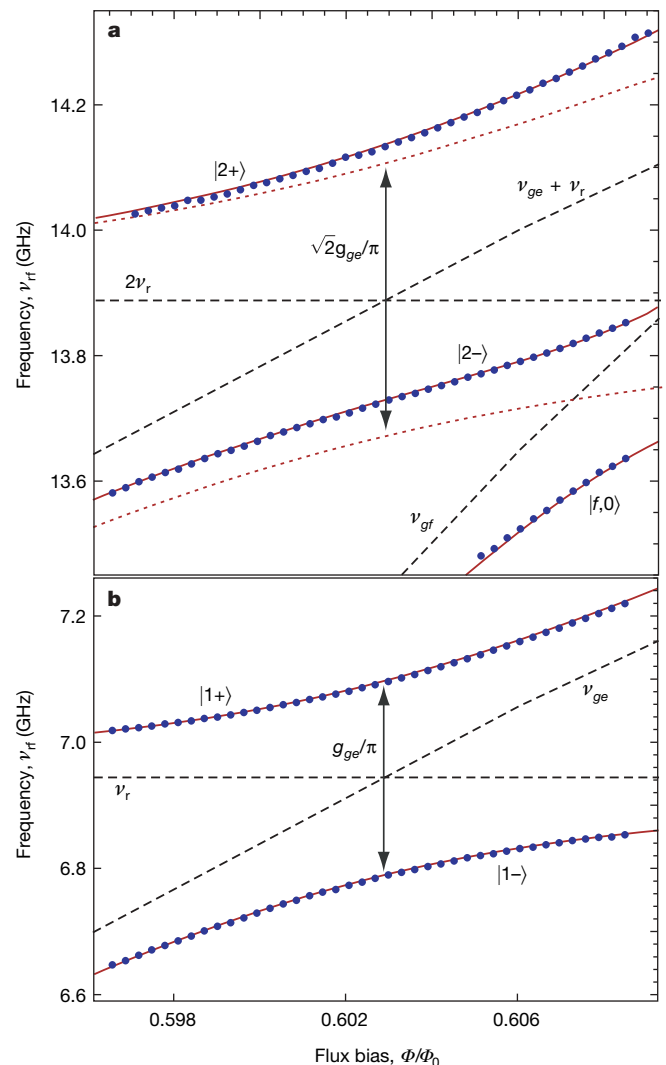
The energies of the first doublet  $|1\pm\rangle$ , split by  $g_{ge}/\pi$  on resonance, are in excellent agreement with the dressed states theory (solid red lines) over the full range of flux-controlled detunings (Fig. 5). The absolute energies of the second doublet states  $|2\pm\rangle$  are obtained by adding the extracted probe tone frequencies  $\nu_{1-,2-}$  and  $\nu_{1+,2+}$  to the applied pump frequencies  $\nu_{g0,1-}$  or  $\nu_{g0,1+}$  (blue dots in Fig. 5). For the second doublet, we observe two peaks split by  $1.34g_{ge}/\pi$  on resonance, a value very close to the expected  $\sqrt{2}\approx 1.41$ . This small frequency shift can easily be understood, without any fitting parameters, by taking into account a third qubit level  $|f, 0\rangle$ , which is at frequency  $\nu_{gf}\approx 2\nu_{ge}-E_C$  for the transmon type qubit<sup>25</sup>, just below the second doublet states  $|2\pm\rangle$ . In order to find the energies of the dressed states in the presence of this additional level, we diagonalize the hamiltonian

$\hat{\mathcal{H}}=\hat{\mathcal{H}}_0+\hat{\mathcal{H}}_1$ , where  $\hat{\mathcal{H}}_1=\hbar\omega_{gf}\hat{\sigma}_{ff}+g_{gf}(\hat{\sigma}_{ef}^\dagger a+a^\dagger\hat{\sigma}_{ef})$  and  $g_{ef}/2\pi\approx 210$  MHz (obtained from exact diagonalization) denotes the coupling of the  $|e\rangle$  to  $|f\rangle$  transition to the cavity. The presence of the  $|f, 0\rangle$  level is observed to shift the antisymmetric state  $|2-\rangle$ , being closer in frequency to the  $|f, 0\rangle$  state, more than the symmetric state  $|2+\rangle$  (Figs 1 and 5), leading to the small difference of the observed splitting from  $\sqrt{2}$ . The  $|f, 0\rangle$  state, being dressed by the states  $|g, 2\rangle$  and  $|e, 1\rangle$ , is also directly observed in the spectrum via the transition  $|1-\rangle \rightarrow |f, 0\rangle$  at frequency  $\nu_{1-,f0}$  (Fig. 4c). This is in excellent agreement with the dressed states model (Fig. 5). For comparison, the dressed states split by  $\sqrt{2}g_{ge}/\pi$  in the absence of the  $|f, 0\rangle$  state are shown as dotted red lines in Fig. 5.

Our experiments clearly demonstrate the quantum nonlinearity of a system of one or two photons strongly coupled to a single artificial atom in a cavity QED setting. Both symmetric and antisymmetric superposition states involving up to two photons are resolved by many tens of linewidths. Recently, signatures of the  $|2-\rangle$  state have also been observed spectroscopically in an independent work on optical cavity QED<sup>30</sup>. We have also observed that higher excited states of the artificial atom can induce energy shifts in the coupled



**Figure 4 | Vacuum Rabi mode splitting with two photons.** **a**, Cavity transmission  $T$  as in Fig. 3 with an additional pump tone applied to the resonator input at frequency  $\nu_{g0,1+}$  populating the  $|1+\rangle$  state. **b**, Spectrum at  $\Delta = 0$ , indicated by arrows in **a**. **c**, Transmission  $T$  with a pump tone applied at  $\nu_{g0,1-}$  populating the  $|1-\rangle$  state. **d**, Spectrum at  $\Phi/\Phi_0 \approx 0.606$ , indicated by arrows in **c**. See text for details of pump tone nomenclature.



**Figure 5 | Experimental dressed state energy levels.** Measured dressed state energies (blue dots) reconstructed by summing pump and probe frequencies, compared to the calculated uncoupled cavity and qubit levels (dashed lines), the calculated dressed state energies in the qubit two-level approximation (dotted) and to the corresponding calculation including the third qubit level (solid red lines). Panels show  $\nu_{rf}$  ranges around  $2\nu_r$  (**a**) and  $\nu_r$  (**b**).



atom–photon states. These shifts should also be observable in time resolved measurements of Rabi oscillations with photon number states. In our circuit QED system, excited states  $|n\pm\rangle$  with  $n > 2$  are also observable (not shown) both by pumping the system with thermal photons and by applying strong coherent drive fields inducing multi-photon transitions. The observed very strong nonlinearity on the level of a single quantum (or a few quanta) could be used for the realization of a single-photon transistor, parametric down-conversion, and for the generation and detection of individual microwave photons.

Received 18 March; accepted 20 May 2008.

- Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).
- Mabuchi, H. & Doherty, A. C. Cavity quantum electrodynamics: Coherence in context. *Science* **298**, 1372–1377 (2002).
- Walther, H., Varcoe, B. T. H., Englert, B.-G. & Becker, T. Cavity quantum electrodynamics. *Rep. Prog. Phys.* **69**, 1325–1382 (2006).
- Reithmaier, J. P. *et al.* Strong coupling in a single quantum dot-semiconductor microcavity system. *Nature* **432**, 197–200 (2004).
- Yoshie, T. *et al.* Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432**, 200–203 (2004).
- Peter, E. *et al.* Exciton-photon strong-coupling regime for a single quantum dot embedded in a microcavity. *Phys. Rev. Lett.* **95**, 067401 (2005).
- Hennessy, K. *et al.* Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445**, 896–899 (2007).
- Englund, D. *et al.* Controlling cavity reflectivity with a single quantum dot. *Nature* **450**, 857–861 (2007).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Chiorescu, I. *et al.* Coherent dynamics of a flux qubit coupled to a harmonic oscillator. *Nature* **431**, 159–162 (2004).
- Johansson, J. *et al.* Vacuum Rabi oscillations in a macroscopic superconducting qubit LC oscillator system. *Phys. Rev. Lett.* **96**, 127006 (2006).
- Zhu, Y. *et al.* Vacuum Rabi splitting as a feature of linear-dispersion theory: Analysis and experimental observations. *Phys. Rev. Lett.* **64**, 2499–2502 (1990).
- Walls, D. & Milburn, G. *Quantum Optics* (Springer, Berlin, 1994).
- Carmichael, H. J., Kochan, P. & Sanders, B. C. Photon correlation spectroscopy. *Phys. Rev. Lett.* **77**, 631–634 (1996).
- Blais, A., Huang, R. S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A* **69**, 062320 (2004).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Thompson, R. J., Rempe, G. & Kimble, H. J. Observation of normal-mode splitting for an atom in an optical cavity. *Phys. Rev. Lett.* **68**, 1132–1135 (1992).
- Boca, A. *et al.* Observation of the vacuum Rabi spectrum for one trapped atom. *Phys. Rev. Lett.* **93**, 233603 (2004).
- Brune, M. *et al.* Quantum Rabi oscillation: A direct test of field quantization in a cavity. *Phys. Rev. Lett.* **76**, 1800–1803 (1996).
- Varcoe, B. T. H., Brattke, S., Weidinger, M. & Walther, H. Preparing pure photon number states of the radiation field. *Nature* **403**, 743–746 (2000).
- Bertet, P. *et al.* Generating and probing a two-photon Fock state with a single atom in a cavity. *Phys. Rev. Lett.* **88**, 143601 (2002).
- Thompson, R. J., Turchette, Q. A., Carnal, O. & Kimble, H. J. Nonlinear spectroscopy in the strong-coupling regime of cavity QED. *Phys. Rev. A* **57**, 3084–3104 (1998).
- Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
- Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
- Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
- Schreier, J. A. *et al.* Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B* **77**, 180502(R) (2008).
- Bouchiat, V., Vion, D., Joyez, P., Esteve, D. & Devoret, M. H. Quantum coherence with a single Cooper pair. *Phys. Scripta* **T76**, 165–170 (1998).
- Schuster, D. I. *et al.* AC Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
- Schoelkopf, R. J. & Girvin, S. M. Wiring up quantum systems. *Nature* **451**, 664–669 (2008).
- Schuster, I. *et al.* Nonlinear spectroscopy of photons bound to one atom. *Nature Phys.* **4**, 382–385 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. S. Bishop, J. M. Chow, T. Esslinger, L. Frunzio, A. Imamoglu, B. R. Johnson, J. Koch, R. J. Schoelkopf and D. I. Schuster for discussions. This work was supported by SNF and ETHZ. P.J.L. was supported by the EU with an MC-EIF. A.B. was supported by NSERC, CIFAR and FQRNT.

**Author Contributions** J.M.F. performed the experiments and analysed the data using theory developed by A.B.; M.G. designed and fabricated the sample; M.B. contributed to sample characterization; R.B. contributed to the realization of the experimental set-up; and J.M.F. and A.W. co-wrote the paper. All authors discussed the results and commented on the manuscript. P.J.L. and A.W. supervised this work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.M.F. (jfink@phys.ethz.ch) or A.W. (andreas.wallraff@phys.ethz.ch).

# Imaging and dynamics of light atoms and molecules on graphene

Jannik C. Meyer<sup>1,2</sup>, C. O. Girit<sup>1,2</sup>, M. F. Crommie<sup>1,2</sup> & A. Zettl<sup>1,2</sup>

Observing the individual building blocks of matter is one of the primary goals of microscopy. The invention of the scanning tunnelling microscope<sup>1</sup> revolutionized experimental surface science in that atomic-scale features on a solid-state surface could finally be readily imaged. However, scanning tunnelling microscopy has limited applicability due to restrictions in, for example, sample conductivity, cleanliness, and data acquisition rate. An older microscopy technique, that of transmission electron microscopy (TEM)<sup>2,3</sup>, has benefited tremendously in recent years from subtle instrumentation advances, and individual heavy (high-atomic-number) atoms can now be detected by TEM<sup>4–7</sup> even when embedded within a semiconductor material<sup>8,9</sup>. But detecting an individual low-atomic-number atom, for example carbon or even hydrogen, is still extremely challenging, if not impossible, via conventional TEM owing to the very low contrast of light elements<sup>2,3,10–12</sup>. Here we demonstrate a means to observe, by conventional TEM, even the smallest atoms and molecules: on a clean single-layer graphene membrane, adsorbates such as atomic hydrogen and carbon can be seen as if they were suspended in free space. We directly image such individual adatoms, along with carbon chains and vacancies, and investigate their dynamics in real time. These techniques open a way to reveal dynamics of more complex chemical reactions or identify the atomic-scale structure of unknown adsorbates. In addition, the study of atomic-scale defects in graphene may provide insights for nanoelectronic applications of this interesting material.

The atomic-scale resolution of TEM comes at the price of requiring that the transmitted electron beam reach the imaging lenses and detector, and so TEM works only for ultra-thin, electron-transparent samples. In high-resolution TEM and all related techniques—such as electron diffraction, scanning transmission electron microscopy, electron energy loss spectroscopy or elemental mapping—any support film or membrane provides a background signal that is most significant for the smallest objects under investigation. Individual nanoscale particles or molecules usually need to be supported by a continuous membrane, as only tubular or rod-shaped nanoparticles (such as carbon nanotubes) can be suspended across holes in the membrane. Indeed, single-walled carbon nanotubes have been used for low-background TEM studies of encapsulated molecules<sup>13–15</sup> or of defects in the cylinder-shaped graphene sheets<sup>16,17</sup>. However, the limited space, harsh filling procedures, and strongly curved shape of the sheet limit the applicability and complicate the analysis.

As we demonstrate below, a graphene membrane provides the ultimate sample support for electron microscopy. With a thickness of only one atom, it is the thinnest possible continuous material. Owing to its crystalline nature, a graphene support membrane is either completely invisible or, if the graphene lattice is resolved by a very-high-resolution microscope, its contribution to the imaging

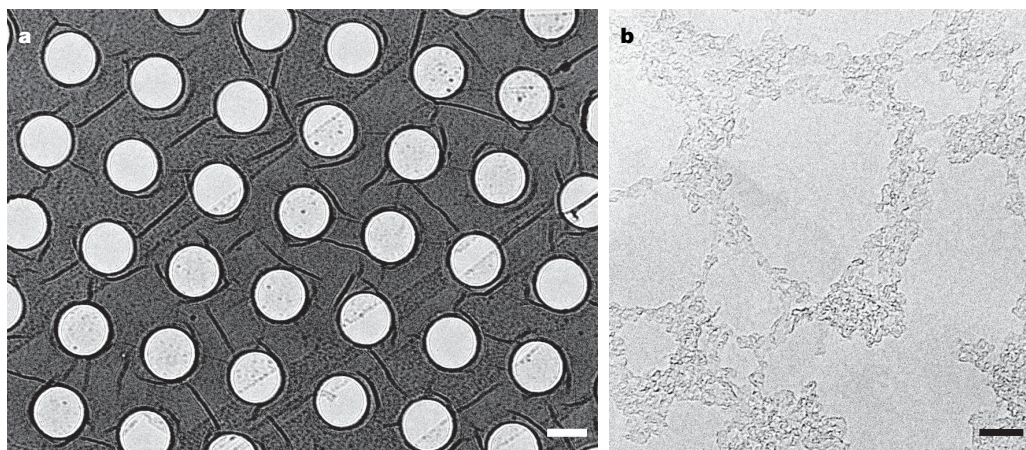
signal can be easily subtracted. Graphene is also a good electrical conductor and therefore displays minimal charging effects from the electron beam. Remarkably, we find that a graphene membrane enables single-adatom sensitivity even when using a common TEM that does not resolve a graphitic lattice.

In order to observe adsorbates at the single-atom level, the graphene support membrane must be exceptionally clean. In contrast to an earlier graphene membrane preparation method<sup>18</sup>, our approach does not rely on electron beam lithography and is simple enough to be reproducible in any basic microscopy laboratory (a detailed description of the sample preparation is given in the Supplementary Information). In brief, we start with graphene prepared by mechanical cleavage with an adhesive tape<sup>19–21</sup>, and transfer selected sheets to commercially available TEM grids. We use electron diffraction analysis<sup>18</sup> to verify the presence of a single layer. Figure 1a shows a low-magnification view of a graphene sheet suspended across the 1.3  $\mu\text{m}$  holes of the perforated carbon foil, with a close-up shown in Fig. 1b. More than 50% of the area on these graphene membranes appears exceptionally clean, with no dramatic contrast in high-resolution TEM images (Fig. 1b). As we now demonstrate, however, these ‘clean’ regions contain individual adatoms that are readily observable by TEM. Although individual exposures can reveal useful data, a dramatic improvement in the signal-to-noise ratio is achieved by summing multiple subsequent frames (corrected for sample drift), which effectively increases the exposure time beyond the dynamic range of the TEM CCD detector. Summing as few as five frames yields striking visual improvement with atomic-scale features (including individual adatoms) becoming readily apparent, and summing 100 frames reduces the noise to below 0.12% (standard deviation in a relatively featureless region of the graphene membrane).

Figure 2a shows a TEM image in which an individual carbon atom, attached to the graphene membrane, is identified by an arrow. We recorded eight consecutive images essentially identical to that of Fig. 2a (each a summation of 20 frames on the CCD), demonstrating that the carbon atom did not adsorb or desorb during the time of exposure. To identify the adatom, image simulations were carried out as described in the Supplementary Information. The good agreement between the TEM data and image simulation (Fig. 2b, c) confirms the carbon atom identification. Simulations for individual boron, nitrogen or oxygen adatoms also provide reasonable fits; however, carbon is the dominant component of vacuum contamination and surface adsorbates within our TEM, making these other candidates unlikely.

Closer inspection of Fig. 2a also reveals faint atomic-scale structure distinctly different from carbon adatoms. To highlight these faint features, we show in Fig. 2d a summation of 100 consecutive TEM frames for the same physical region. Figure 2d reveals a moderate density of additional dark features (dark grey points, a selection of

<sup>1</sup>Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. <sup>2</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.



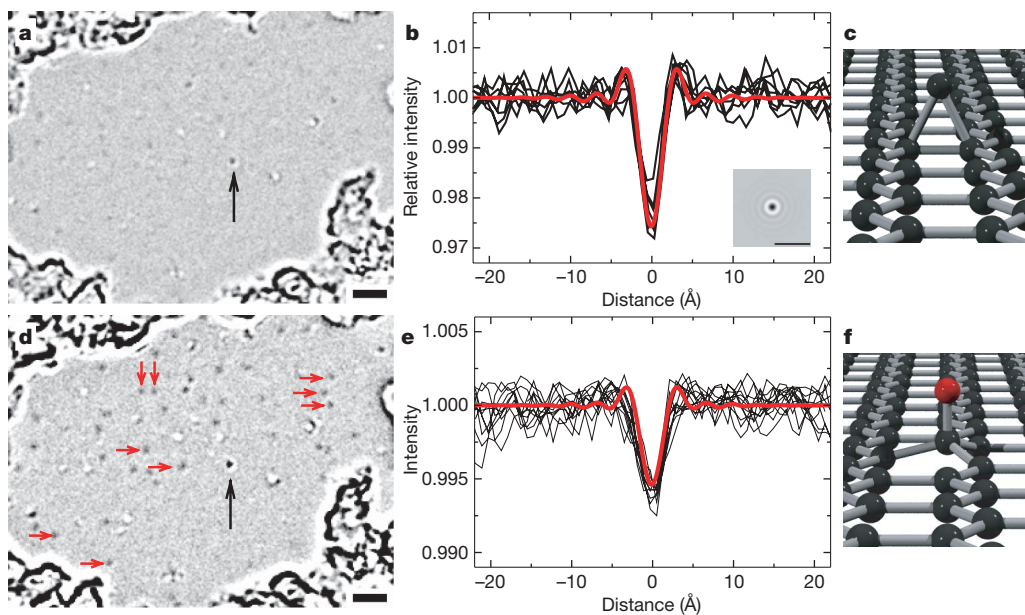
**Figure 1 | Graphene membrane sample as observed by TEM.** **a**, Low-magnification overview image of a suspended graphene sheet on the perforated carbon foil. **b**, High-resolution close-up of a graphene membrane.

We observe small, extremely clean areas with diameters of 10–50 nm where no contrast is visible, separated by regions with thin amorphous adsorbates. Scale bars; 1  $\mu\text{m}$  (**a**), 10 nm (**b**).

which are identified with red arrows) with identical intensity profiles, all with a central dip reduction near 0.6% of the mean bright-field intensity (Fig. 2e). By comparing the TEM image data for these additional features to adatom simulations, we rule out any adatom heavier than helium, as well as a substitution of carbon atoms in the graphene membrane by other elements. However, a hydrogen adatom results in the correct 0.6% dip in the bright-field intensity, shown by the red curve in Fig. 2e. The large number of essentially identical adatom profiles, along with the excellent agreement with the simulated contrast, provides convincing evidence that we have, for the first time to our knowledge, detected individual hydrogen atoms by TEM.

We emphasize that we do not claim to have resolved a hydrogen–carbon distance, which would require advanced aberration-corrected instrumentation. However, detecting an isolated hydrogen atom against a nearly invisible background only requires an adequate signal-to-noise ratio. Electron scattering from hydrogen has been detected previously in electron diffraction experiments<sup>22</sup>, and was

found to produce a three to four times lower signal than carbon, in agreement with our values. In order to verify the uniqueness of the hydrogen atom identification, we consider whether alternative structures may lead to the observed contrast. The contrast match to simulations, using edges or vacancies in the sheet as reference, is better than a factor of two (see Supplementary Information) in our experiment. For an adatom on the sheet, only hydrogen or helium can produce the observed contrast. We rule out helium, as it will not bind to carbon and is not present anywhere in the experiment. We considered all known defects of graphene sheets. All vacancies, relaxed<sup>23</sup> or not, will produce a white spot. An adatom–vacancy pair<sup>16</sup>, even with the minimum separation below the resolution of our microscope, would show a white-and-black symmetric intensity at a detectable level according to simulations, and was indeed observed. A Stone–Wales defect<sup>23</sup> is also expected to have a white-and-black symmetric contrast with zero mean value. Extended defects, such as dislocations, would not produce the rotationally symmetric contrast of an adatom. Finally, a single-layer graphene



**Figure 2 | Adatom images.** **a**, Carbon adatom (indicated by black arrow). **b**, Intensity profiles from several images of the carbon adatom (black traces), and a simulated profile (red trace). Inset, simulated image. **c**, Carbon adatom configuration according to ref. 30. **d**, Hydrogen adatoms on the same sample (dark grey spots), a selection of which are indicated by red arrows. Black

arrow again indicates the carbon adatom. **e**, Profile plots of selected hydrogen adatoms from **d** (black traces). Red line is the simulated profile for a hydrogen adatom. **f**, Configuration of a chemisorbed hydrogen atom (red) according to ref. 25. All scale bars, 2 nm.



membrane has a much smaller set of possible defects than graphite (or few-layer graphene): for example, it can not have interstitials or bonds between layers. Thus, we confirm the identification of the hydrogen adatom.

In addition to individual adatoms, we observe by the same TEM imaging methods the generation (by the electron beam) and dynamics of defects (vacancies) in the graphene membrane, as well as the dynamics of a variety of molecular-scale adsorbates. The formation of vacancies owing to knock-on damage by the electron beam is shown in Fig. 3a–c. We also observe vacancies that disappear by interaction with mobile adsorbates. Larger adsorbates (small molecules) become trapped preferentially at defects, and can be observed at one position for typically one to five minutes. Frequently, we see that the vacancy disappears along with the trapped adsorbate (Fig. 3d–f), and the missing carbon atom has obviously been resubstituted from the adsorbate. Further, we can directly observe linear molecules on graphene membranes (Fig. 4) that resemble an individual alkane or alkene carbon chain. These molecules are found to spontaneously appear in the field of view, presumably adsorbed onto the graphene membrane from the vacuum contamination. We can follow their dynamics for several minutes, as shown in Fig. 4b–d and in the Supplementary Videos.

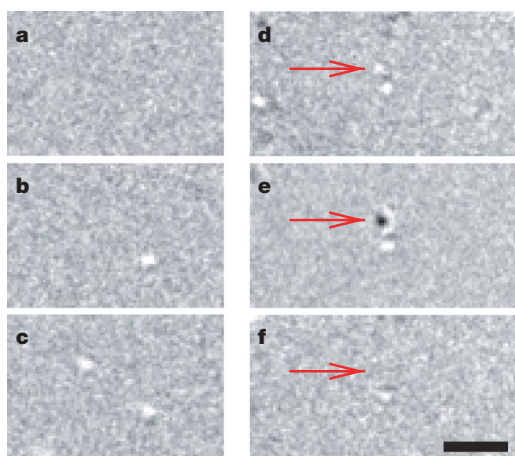
The remarkable TEM imaging capability afforded by a suspended, single graphene membrane warrants further discussion. For an ideal graphene sheet, there are no components in the structure with a period larger than 2.1 Å, which is beyond the information limit of approximately 2.9 Å for the microscope used in the present studies (JEOL 2010 operated at 100 kV). Therefore, although the ideal graphene membrane cannot be resolved under these conditions, any perturbation to the crystalline structure can be detected as long as a sufficient number of electrons can be recorded for statistical significance. Indeed, our graphene membranes are highly stable in the electron beam at 100 kV, allowing long data collection times on one region. For example, all images in Figs 2–4 are recorded from graphene membranes after between one and three hours of irradiation (at  $\sim 7 \text{ A cm}^{-2}$ ). Moreover, the summation of 100 consecutive CCD frames corresponds to an exposure time of 20 min, and distortions in the membrane during this time are below the resolution limit. This combination of a crystalline, atomically thin membrane along with the high beam stability and the absence of an amorphous background signal on the nominally clean membrane enables this unprecedented single-light-atom sensitivity in TEM. In comparison, single-walled

carbon nanotubes show strong deformations under the same dose and energy of electron irradiation (see figure 5 of ref. 24), probably because the cylindrical geometry allows beam-induced defects to relax via local deformations more easily.

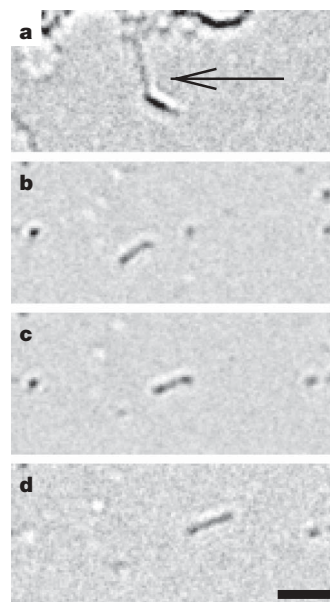
The observation of stable and well-localized hydrogen adatoms on graphene, in spite of the irradiation and room temperature conditions, implies that these are chemisorbed rather than physisorbed atoms. Strong bonding of hydrogen to graphite is possible if the nearest carbon atom changes its bonds from  $sp^2$  to  $sp^3$  configuration<sup>25–27</sup>, with the carbon atom displaced from the plane by about 0.36 Å (Fig. 2f). Moreover, it was found<sup>25</sup> that hydrogen cannot bind to graphene if the carbon is confined to a plane (for example by strong bonding to a substrate), whereas an isolated membrane can deform easily to accommodate different types of bonds<sup>28,29</sup>. From the observed density of hydrogen adatoms, we conclude that only about 0.3% of the carbon atoms in our graphene membrane are in an  $sp^3$  configuration with a hydrogen adatom.

Our real-time observation of molecular dynamics has important implications for chemical diffusion and reaction dynamics studies. As demonstrated above, a variety of molecular-scale adsorbates become trapped on the membrane, and often detach again or decompose after a few minutes. We can observe individual alkane-type molecules and we can even follow their migration. Alkanes are an essential ingredient of organic chemistry, and it therefore appears likely that other, more complex, molecules could be observed after deposition on graphene membranes. We find that the carbon chains are sufficiently stable and localized for characterization even at room temperature, and note that these adsorbates were only trapped on the membrane after a moderate density of defects had been created by irradiation.

In conclusion, we have demonstrated that graphene membranes enable TEM visualization of ultra-low-contrast objects. The imaging of individual hydrogen and carbon adatoms and carbon chains demonstrates a new level of sensitivity that is relevant for organic materials. A key strength of the TEM is its ability to image individual entities rather than averaging over an ensemble, and direct imaging promises insights ranging from the characterization of complex chemicals and nanomaterials to biological molecules. The extremely high sensitivity that a graphene membrane in the TEM provides with



**Figure 3 | Dynamics of defects.** a–c, Generation of vacancies owing to electron irradiation. Time between a and c is 50 min. d–f, Annealing of a vacancy by interaction with an adsorbate. We observe two individual vacancies (d), and then trapping of a larger adsorbate (e), corresponding to a mass of a few carbon atoms, on one of the defects. After  $\sim 5$  min, both the adsorbate and one vacancy disappear (f), showing that the missing carbon atom in the graphene sheet has been replaced by an atom from the adsorbate. Scale bar, 2 nm.



**Figure 4 | Molecular scale adsorbates.** a, Molecule suspended between other adsorbates (arrow). b–d, Migration of a carbon chain, where one end remains attached in each step. This migration is also shown in the Supplementary Video. The contrast is in agreement with an alkane molecule. Scale bar, 2 nm.

respect to adsorbates has allowed us to detect even hydrogen, demonstrating the ultimate in TEM atomic sensitivity. Whereas the study of defects, vacancies and edges of the graphene sheet itself will provide insights for potential electronic modifications of this new material, the placement of objects on graphene membranes will enable unprecedented analysis by TEM, including electron spectroscopic analysis, and the study of molecular dynamics.

## METHODS SUMMARY

Graphene sheets are prepared on oxidized silicon substrates by mechanical cleavage. After identification by optical microscopy, selected sheets are transferred to Quantifoil TEM grids (Quantifoil Micro Tools) with 1.2 µm holes. The perforated TEM support film is brought into contact with the substrate and graphene sheet by evaporating a drop of solvent. The substrate is then removed by wet chemistry, while the graphene sheet remains attached to the TEM grids (details are given in the Supplementary Information). TEM imaging is carried out in a JEOL 2010 microscope operated at 100 kV. The sample holder is at room temperature; the actual sample temperature may differ owing to electron beam heating or the nearby decontaminator cold trap. A continuous sequence of images is recorded on the CCD camera. The defocus value (60 nm) and presence of vibrations is estimated from the thin amorphous coverage that intersperses the clean areas of the graphene membrane for each frame, and ~5% of the frames are discarded. Then, drift-compensated summations of up to 100 frames are performed (with each frame verified for imaging parameters and vibrations) to obtain an adequate signal-to-noise ratio. Orthogonal slices through the stack of images (see Supplementary Information) clearly establish whether a feature of interest has been static during the entire effective exposure time, or can be used to detect interesting dynamics in the data.

Received 10 December 2007; accepted 7 May 2008.

1. Binnig, G., Rohrer, H., Gerber, Ch. & Weibel, E. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* **49**, 57–61 (1982).
2. Spence, J. C. H. *High-Resolution Electron Microscopy* (Oxford Univ. Press, Oxford, UK, 2003).
3. Buseck, P. R., Cowley, J. M. & Eyring, L. *High-Resolution Transmission Electron Microscopy* (Oxford Univ. Press, Oxford, UK, 1988).
4. Crewe, A. V., Wall, J. & Langmore, J. Visibility of single atoms. *Science* **168**, 1338–1340 (1970).
5. Hashimoto, H. *et al.* Visualization of single atoms in molecules and crystals by dark field electron microscopy. *J. Electron Microsc. (Tokyo)* **22**, 123–134 (1973).
6. Iijima, S. Observation of single and clusters of atoms in bright field electron microscopy. *Optik* **48**, 193–213 (1977).
7. Nellist, P. D. & Pennycook, S. J. Direct imaging of the atomic configuration of ultradispersed catalysts. *Science* **274**, 413–415 (1996).
8. Voyles, P. M., Muller, D. A., Graul, J. L., Citrin, P. H. & Gossman, H.-J. L. Atomic-scale imaging of individual dopant atoms and clusters in highly n-type bulk silicon. *Nature* **416**, 826–829 (2002).
9. van Benthem, K. *et al.* Three-dimensional imaging of individual hafnium atoms inside a semiconductor device. *Appl. Phys. Lett.* **87**, 034104 (2005).
10. Doyle, P. A. & Turner, P. S. Relativistic Hartree-Fock x-ray and electron scattering factors. *Acta Crystallogr. A* **24**, 390–397 (1968).
11. Kisielowski, C. *et al.* Imaging columns of the light elements carbon, nitrogen and oxygen with sub Angstrom resolution. *Ultramicroscopy* **89**, 243–263 (2001).
12. Jia, C. L., Lentzen, M. & Urban, K. Atomic-resolution imaging of oxygen in perovskite ceramics. *Science* **299**, 870–873 (2003).
13. Smith, B. W., Monthieux, M. & Luzzi, D. E. Encapsulated C in carbon nanotubes. *Nature* **396**, 323–324 (1998).
14. Liu, Z. *et al.* Transmission electron microscopy imaging of individual functional groups of fullerene derivatives. *Phys. Rev. Lett.* **96**, 088304 (2006).
15. Lui, Z., Yanagi, K., Suenaga, K., Kataura, H. & Iijima, S. Imaging the dynamic behaviour of individual retinal chromophores confined inside carbon nanotubes. *Nature Nanotechnol.* **2**, 422–425 (2007).
16. Hashimoto, A., Suenaga, K., Gloter, A., Urita, K. & Iijima, S. Direct evidence for atomic defects in graphene layers. *Nature* **430**, 870–873 (2004).
17. Suenaga, K. *et al.* Imaging active topological defects in carbon nanotubes. *Nature Nanotechnol.* **2**, 358–360 (2007).
18. Meyer, J. C. *et al.* The structure of suspended graphene sheets. *Nature* **446**, 60–63 (2007).
19. Novoselov, K. S. *et al.* Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).
20. Novoselov, K. S. *et al.* Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
21. Zhang, Y., Tan, J. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
22. Vainshtein, B. K. & Pinsker, Z. G. Opređenje Polozheniya Vodoroda V Kristallicheskoj Reshetke Parafina. *Dokl. Akad. Nauk SSSR* **72**, 53–56 (1950).
23. Amara, H., Latil, S., Lambin, Ph. & Charlier, J.-C. Scanning tunneling microscopy fingerprints of point defects in graphene: A theoretical prediction. *Phys. Rev. B* **76**, 115423 (2007).
24. Smith, B. W. & Luzzi, E. Electron irradiation effects in single wall carbon nanotubes. *J. Appl. Phys.* **90**, 3509–3515 (2001).
25. Jeloaica, L. & Sidis, V. DFT investigation of the adsorption of atomic hydrogen on a cluster-model graphite surface. *Chem. Phys. Lett.* **300**, 157–162 (1999).
26. Sha, X. & Jackson, B. First-principles study of the structural and energetic properties of H atoms on a graphite (0001) surface. *Surf. Sci.* **496**, 318–330 (2002).
27. Hornekaer, L. *et al.* Metastable structures and recombination pathways for atomic hydrogen on the graphite (0001) surface. *Phys. Rev. Lett.* **96**, 156104 (2006).
28. Boukhvalov, D. W., Katsnelson, M. I. & Lichtenstein, A. I. Hydrogen on graphene: Electronic structure, total energy, structural distortions and magnetism from first-principles calculations. *Phys. Rev. B* **77**, 035427 (2007).
29. Ito, A., Nakamura, H. & Takayama, A. Chemical reaction between single hydrogen atom and graphene. Preprint at (<http://arxiv.org/abs/cond-mat/0703377>) 2007.
30. Nordlund, K., Keinonen, J. & Mattila, T. Formation of ion irradiation induced small-scale defects on graphite surfaces. *Phys. Rev. Lett.* **77**, 699–702 (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, of the US Department of Energy under contract DE-AC02-05CH11231. A.Z. acknowledges support from the Miller Institute of Basic Research in Science, and C.O.G. acknowledges support from an NSF Graduate Fellowship.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.C.M. (email:[jannikmeyer.de](mailto:jannikmeyer.de)) or A.Z. ([azettl@berkeley.edu](mailto:azettl@berkeley.edu)).

# Cretaceous oceanic anoxic event 2 triggered by a massive magmatic episode

Steven C. Turgeon<sup>1</sup> & Robert A. Creaser<sup>1</sup>

Oceanic anoxic events (OAEs) were episodes of widespread marine anoxia during which large amounts of organic carbon were buried on the ocean floor under oxygen-deficient bottom waters<sup>1,2</sup>. OAE2, occurring at the Cenomanian/Turonian boundary (about 93.5 Myr ago)<sup>3</sup>, is the most widespread and best defined OAE of the mid-Cretaceous. Although the enhanced burial of organic matter can be explained either through increased primary productivity or enhanced preservation scenarios<sup>1,2</sup>, the actual trigger mechanism, corresponding closely to the onset of these episodes of increased carbon sequestration, has not been clearly identified. It has been postulated that large-scale magmatic activity initially triggered OAE2 (refs 4, 5), but a direct proxy of magmatism preserved in the sedimentary record coinciding closely with the onset of OAE2 has not yet been found. Here we report seawater osmium isotope ratios in organic-rich sediments from two distant sites. We find that at both study sites the marine osmium isotope record changes abruptly just at or before the onset of OAE2. Using a simple two-component mixing equation, we calculate that over 97 per cent of the total osmium content in contemporaneous seawater at both sites is magmatic in origin, a ~30–50-fold increase relative to pre-OAE conditions. Furthermore, the magmatic osmium isotope signal appears slightly before the OAE2—as indicated by carbon isotope ratios—suggesting a time-lag of up to ~23 kyr between magmatism and the onset of significant organic carbon burial, which may reflect the reaction time of the global ocean system. Our marine osmium isotope data are indicative of a widespread magmatic pulse at the onset of OAE2, which may have triggered the subsequent deposition of large amounts of organic matter.

OAE2, occurring at the Cenomanian/Turonian boundary (~93.5 Myr ago)<sup>3</sup>, is one of several Mesozoic intervals of widespread ocean anoxia<sup>1,2</sup>. OAE2 resulted in a worldwide deposition of organic-rich (black) shales and is demonstrated by an abrupt positive <sup>13</sup>C/<sup>12</sup>C (δ<sup>13</sup>C) excursion<sup>6</sup> in bulk organic matter (up to 6‰) and carbonates (2–3‰)<sup>7</sup>. This isotopic shift, which can be correlated at several distant sites<sup>7</sup>, is widely thought to have resulted from an increase in sedimentary burial of <sup>13</sup>C-depleted organic carbon in response to the anoxic conditions<sup>8</sup>. In addition to a selective extinction most severely affecting deep-sea benthic foraminifera<sup>9</sup>, this episode of carbon sequestration seems to have led to an ephemeral but significant decrease<sup>6</sup> in the partial pressure of atmospheric CO<sub>2</sub> and cooling of surface temperatures<sup>8,10</sup>, making this interval of particular interest for studies of the effects of climate change.

It has often been proposed that an important magmatic event such as the emplacement of large igneous provinces (LIPs) or mid-ocean ridges may have triggered OAE2 (refs 4, 5). Although a general consensus on a causal link to the deposition of black shale has yet to be achieved, several feedback scenarios have been proposed. Increased magmatism, for example, in addition to emitting large quantities of CO<sub>2</sub> to the atmosphere, would also have depleted dissolved oxygen

through reactions with trace metals and sulphides in hydrothermal fluids<sup>5</sup>, thereby increasing the preservation of organic matter on the sea floor. Other feedback mechanisms involve enhancing primary productivity in the oceans either through the release of reduced compounds and biologically limiting elements (for example, iron from a hydrothermal-derived plume<sup>4</sup>), or through increased nutrient flux owing to high sea levels engendered by increased seafloor spreading or remineralization of organic matter in the water column<sup>5,11</sup>.

If flood basalt eruption did indeed have a causative role in triggering OAE2, then a major magmatic episode should be coeval with, or should immediately precede, the carbon isotope excursion. However, the linking of magmatism to OAE initiation has relied either on indirect evidence provided by the coincidence of LIP emplacement (such as the Caribbean–Colombian igneous province<sup>12</sup>, the Kerguelen and Ontong Java plateaux or the Madagascar flood basalts<sup>13,14</sup>) and trace metals in black shales<sup>12</sup>, or on <sup>87</sup>Sr/<sup>86</sup>Sr isotopes in the sedimentary record that document the interplay of weathering versus mantle-derived strontium input to the oceans<sup>15</sup>. However, although the strontium isotope record of carbonate sediments close to the Cenomanian/Turonian boundary trends towards less radiogenic values that are indicative of a relative increase in magmatic and hydrothermal input of strontium to the oceans<sup>15</sup>, the long residence times of strontium in oceans (~3 Myr) does not allow precise correlation of magmatism to the onset of OAE2. In addition, lead, which has an oceanic residence time of only a few decades, shows a shift in isotopic ratios near the base of OAE2 towards values of Caribbean–Colombian igneous plateau and Madagascar flood basalts<sup>16</sup>, two of the largest of the ~20 × 10<sup>6</sup> km<sup>3</sup> of intraoceanic, plume-related basalts erupted near the Cenomanian/Turonian boundary<sup>17</sup>.

For this study, <sup>187</sup>Os/<sup>188</sup>Os isotopic analyses were performed on two black-shale intervals, ~5,500 km apart during the OAE2: from Demerara rise adjacent to the east-northeast coast of South America (Ocean Drilling Project (ODP) Site 1260B), located in the southern proto-North Atlantic during the Cenomanian–Turonian, as well as from the Furlo section in central Italy, formerly located in the western Tethys seaway. The laminated sediments of Site 1260B consist of a mixture of terrigenous detrital material and carbonates, along with large amounts of organic matter (up to 23 wt%). The <sup>13</sup>C/<sup>12</sup>C isotope excursion at this site begins at 426.38 m composite depth (mcd), and the entire OAE2 here is ~1.2 m thick<sup>10</sup> (Fig. 1a). The Furlo section contains several tens of metres of biosiliceous limestones of the Scaglia Bianca Formation, with a ~1-m-thick interval of shales rich in radiolarians and organic carbon (C<sub>org</sub>; up to 16 wt%) of the Livello Bonarelli, representing the sedimentary expression of OAE2 (ref. 7). In addition, several minor C<sub>org</sub>-rich intervals ('black levels') are interbedded within the Scaglia Bianca up to ~20 m beneath the Bonarelli. Although the <sup>13</sup>C/<sup>12</sup>C isotope record here is discontinuous, the excursion seems to increase suddenly close to the base of the Livello Bonarelli (Fig. 1b)<sup>11</sup>.

<sup>1</sup>Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta, T2G 2E3, Canada.



Most of the osmium in black shales is derived from seawater, with the  $^{187}\text{Os}/^{188}\text{Os}$  ratio being representative of contemporaneous seawater<sup>18</sup>. The isotopic composition of dissolved osmium in modern oceans is nearly homogeneous, with a present-day ratio of  $\sim 1.06$  (ref. 18). This ratio reflects mass balance between osmium input from hydrothermal alteration of juvenile oceanic crust ( $^{187}\text{Os}/^{188}\text{Os} = 0.127$ ), average riverine input from continental weathering ( $^{187}\text{Os}/^{188}\text{Os} = 1.4$ ), together with a generally minor extraterrestrial contribution from meteorites and cosmic dust particles ( $^{187}\text{Os}/^{188}\text{Os} = 0.127$ ) to the global ocean<sup>19</sup>. The short oceanic residence time of osmium ( $\sim 10,000$  yr; ref. 18) allows short-term (thousands of years) fluctuations in seawater isotopic composition to be precisely recorded in sediments.

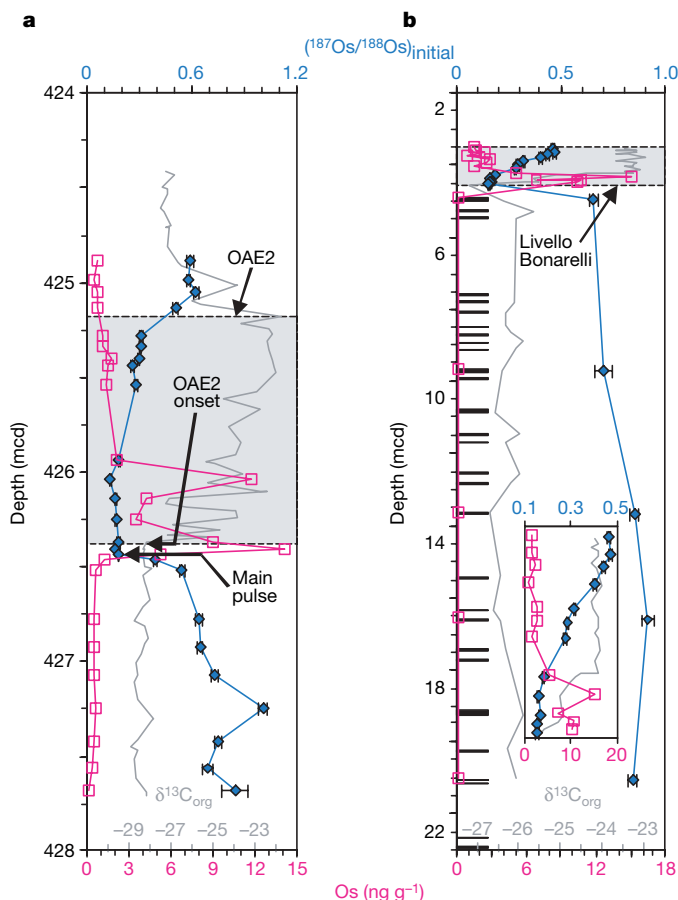
At both study sites, the marine  $^{187}\text{Os}/^{188}\text{Os}$  record changes abruptly just at or before the  $\delta^{13}\text{C}_{\text{org}}$  excursion of  $\sim 3\text{‰}$  (refs 10, 11) defining the OAE2 onset (Fig. 1; see Supplementary Data). At Site 1260B, osmium isotope values decrease irregularly from 0.69–1.0 to 0.54 before OAE2, and decrease to 0.17 a few centimetres before the OAE. These very low osmium isotope values persist through OAE2 to  $\sim 426.0$  mcd before gradually increasing further up the section, and return to  $\sim 0.6$  a few decimetres above OAE2. At Furlo, osmium isotope values in the black levels show a slight decline (from 0.92 to 0.66) beneath the Bonarelli, suggesting that the relative fluxes of the major osmium inputs to oceans remained somewhat steady before the onset of OAE2, perhaps over timescales of 1–3 Myr (ref. 20). An abrupt decrease in marine  $^{187}\text{Os}/^{188}\text{Os}$  from 0.66 to 0.16

occurs between the uppermost black level and the base (comprising roughly the lower one-third) of the Bonarelli. The osmium isotope values then gradually increase further up the section, returning to 0.46. The marine osmium isotope and  $\delta^{13}\text{C}_{\text{org}}$  records are antithetic through the OAE2 at both sites. The  $^{187}\text{Os}/^{188}\text{Os}$  values at the base of OAE2 are similar to the isotopic shift reported for OAE1a (ref. 21), suggesting similar control mechanisms for at least some of the other Cretaceous OAEs. Before OAE2 (up to  $\sim 426.5$  mcd at Site 1260B and pre-Bonarelli at Furlo), osmium contents are low ( $<1$  ng g<sup>-1</sup>) but increase abruptly just before the OAE at Site 1260B and close to the base of the Bonarelli at Furlo to  $>14$  ng g<sup>-1</sup>. In addition, an apparent double spike of osmium enrichment occurs at both sites in the interval of low  $^{187}\text{Os}/^{188}\text{Os}$  ratios (Fig. 1). The similar structure and magnitude of osmium records (contents and isotopic ratios) at both sites strongly suggest that these signals are primary in nature, and that the same process(es) were operating at both sites.

Using a simple two-component mixing equation assuming chondritic (0.127) and bulk continental (1.4) osmium isotopic end-members<sup>18</sup>, we calculate that over 97% of the total osmium content at both sites is magmatic in origin (in comparison with  $\sim 27\%$  for present-day oceans). This represents a  $\sim 30$ – $50$ -fold increase relative to pre-OAE conditions, assuming a constant riverine input. Furthermore, the osmium contents at the base of the OAE2 interval for both sites show enrichment factors of the same order of magnitude as under pre-OAE2 conditions, suggesting that the increased osmium contents mainly reflect a change in the seawater osmium inventory. Osmium isotope data have also been used as a marker of onset of volcanic activity just before the Triassic/Jurassic boundary in the Central Atlantic magmatic province<sup>22</sup> and the Cretaceous/Tertiary boundary in the Deccan flood basalts<sup>23</sup>.

Although the low  $^{187}\text{Os}/^{188}\text{Os}$  values at the OAE might also result from an extraterrestrial source, such as the impact of a large bolide<sup>19</sup>, iridium profiles within the Bonarelli at Furlo<sup>24</sup> do not support this hypothesis, and typical impact-related features such as shocked quartz have not been reported from OAE2. The apparently prolonged (on the order of 190 kyr at Site 1260B, estimated from the duration data of ref. 10) input of mantle-derived osmium with a low  $^{187}\text{Os}/^{188}\text{Os}$  ratio to the oceans is also consistent with existing data showing an increase in several trace elements, including zinc—which is commonly associated with hydrothermal activity<sup>19</sup>—at both Furlo and at Site 1260B (S.C.T., unpublished observations). Our osmium content records at both sites, in conjunction with a marine residence time of osmium in the modern ocean of  $\sim 10$  kyr (ref. 18), indicate a high, but variable, input of osmium to the oceans—probably from at least two major magmatic pulses—during the early phases of OAE2. The specific pathway(s) for transfer of unradiogenic osmium from LIP magmatism to the oceans is poorly constrained, but conceivably involves the seafloor hydrothermal processes invoked for fluxes of other metals at OAE2 (see, for example, ref. 4). However, the osmium contents of both high-temperature and low-temperature hydrothermal fluids from mid-ocean-ridge settings are generally close to, or below, concentrations in sea water<sup>25</sup>. This observation suggests that the release of unradiogenic osmium during LIP emplacement—as indicated by our isotopic data—probably occurred at much higher flux rates and/or fluid concentrations than is typical of known mid-ocean ridge hydrothermal systems.

The difference in timing between the onset of magmatism, demonstrated by the sharp decrease in seawater  $^{187}\text{Os}/^{188}\text{Os}$ , and the carbon isotope excursion (Fig. 1) can be estimated by assuming a constant sediment accumulation rate—a reasonable tenet given that no major trends are observed in lithogenic proxies (for example, titanium, aluminium and rubidium) at either site (ref. 20, and S.C.T., unpublished observations). Consistent with previous work, we use a duration of 550 kyr for the OAE2  $\delta^{13}\text{C}$  excursion, a value in the upper range of proposed durations (ref. 10), and derive an accumulation rate of  $\sim 218$  cm Myr<sup>-1</sup> for OAE2 at Site 1260B. The maximum calculated time-lag between the osmium and  $\delta^{13}\text{C}$  records is  $\sim 23$  kyr;



**Figure 1** | Initial seawater  $^{187}\text{Os}/^{188}\text{Os}$  isotope ratios (blue), Os contents (red), and  $\delta^{13}\text{C}_{\text{org}}$  values (grey) for the two studied sites. **a**, ODP207 Site 1260B ( $n = 25$ ); **b**, Furlo, Italy ( $n = 17$ ). The extent of the OAE for Site 1260B is based on carbon isotope ratios. Horizontal black bars represent location of black levels beneath the Bonarelli at Furlo. The inset shows data points within the Livello Bonarelli. Error bars represent  $2\sigma$  uncertainties of calculated initial seawater osmium isotope ratios.  $\delta^{13}\text{C}_{\text{org}}$  values are from ref. 10 for Site 1260B and from ref. 11 for Furlo.

using an OAE2 duration of only 220 kyr (ref. 26)—a value in the lower range of published duration estimates—the time-lag would be  $\sim 9$  kyr.

Although the exact mechanism responsible for this time-lag is unclear, at Site 1260B a prominent organic-rich layer (426.43–426.38 mcd)—syndepositional with the onset of low osmium ratios—implies an immediate response of the oceans to magmatism, at least at the regional scale. The later shift of  $\delta^{13}\text{C}$  values defining OAE2 onset, however, suggests a feedback mechanism possibly involving a carbon burial ‘threshold’. A similar trend is recorded by sea surface temperatures (ref. 10), which, although seeming to shift before the  $\delta^{13}\text{C}$  excursion, increase only after the onset of magmatism. On this basis, we suggest that the time-lag between osmium and C isotope records resulted from initial buffering of the effects of magmatism by a stratified ocean. It has long been proposed that Cenomanian–Turonian oceans were stratified and characterized by low turnover rates and sluggish circulation<sup>27</sup>. At Site 1260, nanofossils indicative of ocean stratification and a deep nutricline are in decline at the beginning of OAE2, whereas indicators of primary productivity increase only after the  $\delta^{13}\text{C}$  shift<sup>28</sup>. Thus, the initial increase in carbon burial at the onset of OAE2 might have resulted solely from an increased preservation of organic matter related to magmatism-induced bottom-water anoxia. A subsequent breakdown in stratification would have increased surface-water fertility through the shallowing and/or upwelling of nutrient- and biologically limiting element-rich bottom waters, enhancing primary productivity. A scenario of upward expansion of reducing bottom waters during OAE2 is consistent with the observed sequential extinction of benthic foraminifera and, subsequently, intermediate-water foraminifera<sup>29</sup>.

It has long been postulated that large-scale magmatic activity initially triggered the Cenomanian–Turonian OAE, but a direct proxy of magmatism preserved in the sedimentary record, coinciding closely to the onset of OAE2, has not yet been found. The marine osmium isotope record clearly shows that a widespread magmatic pulse occurred just at or before the onset of OAE2 at the Cenomanian/Turonian boundary and probably triggered the subsequent deposition of large amounts of organic matter preserved in the OAE2 interval.

## METHODS SUMMARY

Sediment and a mixed rhenium–osmium spike were digested with a  $\text{CrO}_3$ – $\text{H}_2\text{SO}_4$  solution in sealed tubes. The osmium was then separated with chloroform and subsequently reacted with hydrobromic acid (HBr). The chloroform was then removed and the osmium-bearing HBr was evaporated. The residue was microdistilled twice with a chromium solution to volatilize osmium, and was collected in HBr. Final dry-down of the HBr was performed under a nitrogen atmosphere. An aliquot of the rhenium-bearing  $\text{CrO}_3$ – $\text{H}_2\text{SO}_4$  solution, diluted in water, was reduced by bubbling with  $\text{SO}_2$  cleaned through water in a gas-washing bottle. The rhenium was purified by using anion chromatography. Analytes were then loaded onto nickel (for rhenium) and platinum (for osmium) filaments, and analysed by isotope dilution–thermal ionization mass spectrometry in negative-ion mode. Isotopic ratios were corrected for instrumental mass fractionation, isobaric oxygen interferences, spike and blank contributions. Errors for isotope ratios were determined by numerical error propagation and include uncertainties in spike calibration, weighing, and analytical/instrumental measurements, as well as rhenium bias, oxygen, and blank corrections.

Initial seawater osmium isotopes were determined from

$$(^{187}\text{Os}/^{188}\text{Os})_{\text{initial}} = (^{187}\text{Os}/^{188}\text{Os})_{\text{measured}} - (^{187}\text{Re}/^{188}\text{Os})_{\text{measured}} \times (e^{\lambda t} - 1)$$

where  $\lambda$  is the  $^{187}\text{Re}$  decay constant ( $\lambda^{187}\text{Re}$ ) of  $1.666 \times 10^{-11} \text{ yr}^{-1}$  and  $t$  is the age of the sedimentary rock. The relative mass balance equation used to estimate the contribution of mantle-derived osmium to seawater is  $R_{\text{sw}} = xR_{\text{mtl}} + yR_{\text{chon}} + zR_{\text{cont}}$ , where  $R$  denotes  $^{187}\text{Os}/^{188}\text{Os}$  ratios and the subscript sw represents the calculated initial seawater value, and mtl, chon and cont denote mantle, chondritic and continental sources, respectively. The factors  $x$ ,  $y$  and  $z$  are the relative contributions of each source. Extraterrestrial contributions are assumed

to be negligible<sup>19</sup>. The relative contribution of mantle-derived osmium can therefore be estimated by using  $x = (R_{\text{cont}} - R_{\text{sw}})/(R_{\text{cont}} - R_{\text{mtl}})$ . Values for  $R_{\text{cont}}$  and  $R_{\text{mtl}}$  are 1.4 and 0.127, respectively<sup>18</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 3 October 2007; accepted 1 May 2008.**

- Schlanger, S. O. & Jenkyns, H. C. Cretaceous anoxic events: causes and consequences. *Geol. Mijnb.* **55**, 179–184 (1976).
- Schlanger, S. O., Arthur, M. A., Jenkyns, H. C. & Scholte, P. A. The Cenomanian–Turonian Oceanic Anoxic Event. I. Stratigraphy and distribution of organic carbon-rich beds and the marine  $\delta^{13}\text{C}$  excursion. *Spec. Publ. Geol. Soc. (Lond.)* **26**, 371–399 (1987).
- Gradstein, F. M. *et al.* *A Geologic Time Scale* (Cambridge Univ. Press, Cambridge, 2004).
- Sinton, C. W. & Duncan, R. A. Potential links between ocean plateau volcanism and global ocean anoxia at the Cenomanian–Turonian boundary. *Econ. Geol.* **92**, 836–842 (1997).
- Kerr, A. C. Oceanic plateau formation: a cause of mass extinction and black shale deposition around the Cenomanian–Turonian boundary. *J. Geol. Soc. Lond.* **155**, 619–626 (1998).
- Kuypers, M. M. M., Pancost, R. D. & Sinninghe Damsté, J. S. A large and abrupt fall in atmospheric  $\text{CO}_2$  concentration during Cretaceous times. *Nature* **399**, 342–345 (1999).
- Tsikos, H. *et al.* Carbon-isotope stratigraphy recorded by the Cenomanian–Turonian Oceanic Anoxic Event: correlation and implications based on three key localities. *J. Geol. Soc. Lond.* **161**, 711–719 (2004).
- Arthur, M. A., Dean, W. E. & Pratt, L. M. Geochemical and climatic effects of increased marine organic carbon burial at the Cenomanian/Turonian boundary. *Nature* **335**, 714–717 (1988).
- Kaiho, K. & Hasegawa, T. End-Cenomanian benthic foraminiferal extinctions and oceanic dysoxic events in the northwestern Pacific Ocean. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **111**, 29–43 (1994).
- Forster, A., Schouten, S., Moriya, K., Wilson, P. A. & Sinninghe Damsté, J. S. Tropical warming and intermittent cooling during the Cenomanian/Turonian oceanic anoxic event 2: Sea surface temperature records from the equatorial Atlantic. *Paleoceanography* **22**, doi:10.1029/2006PA001349 (2007).
- Jenkyns, H. C., Matthews, A., Tsikos, H. & Erel, Y. Nitrate reduction, sulfate reduction, and sedimentary iron isotope evolution during the Cenomanian–Turonian oceanic anoxic event. *Paleoceanography* **22**, doi:10.1029/2006PA001355 (2007).
- Snow, L. J., Duncan, R. A. & Bralower, T. J. Trace element abundances in the Rock Canyon Anticline, Pueblo, Colorado, marine sedimentary section and their relationship to Caribbean plateau construction and oxygen anoxic event 2. *Paleoceanography* **20**, doi:10.1029/2004PA001093 (2005).
- Storey, M. *et al.* Timing of hot-spot related volcanism and the breakup of Madagascar and India. *Science* **267**, 852–855 (1995).
- Torsvik, T. H. *et al.* Late Cretaceous magmatism in Madagascar: palaeomagnetic evidence for a stationary Marion hotspot. *Earth Planet. Sci. Lett.* **164**, 221–232 (1998).
- McArthur, J. M., Kennedy, W. J., Chen, M., Thirlwall, M. F. & Gale, A. S. Strontium isotope stratigraphy for Late Cretaceous time: Direct numerical calibration of the Sr isotope curve based on the US Western Interior. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **108**, 95–119 (1994).
- Kuroda, J. *et al.* Contemporaneous massive subaerial volcanism and late Cretaceous Oceanic Anoxic Event 2. *Earth Planet. Sci. Lett.* **256**, 211–223 (2007).
- Wignall, P. B. Large igneous provinces and mass extinctions. *Earth Sci. Rev.* **53**, 1–33 (2001).
- Peucker-Ehrenbrink, B. & Ravizza, G. The marine osmium isotope record. *Terra Nova* **12**, 205–219 (2000).
- Pegram, W. J. & Turekian, K. K. The osmium isotopic composition change of Cenozoic sea water as inferred from a deep-sea core corrected for meteoritic contributions. *Geochim. Cosmochim. Acta* **63**, 4053–4058 (1999).
- Turgeon, S. & Brumsack, H.-J. Anoxic vs dysoxic events reflected in sediment geochemistry during the Cenomanian–Turonian Boundary Event (Cretaceous) in the Umbria–Marche Basin of central Italy. *Chem. Geol.* **234**, 321–339 (2006).
- Suzuki, K. *et al.* Secular change of Early Cretaceous seawater Os isotope composition: an indicator of a LIP–OAE link (abstract). *Japan Geoscience Union Meeting*, 19–24 May 2007, <[www.jpгу.org/publication/cd-rom/2007cd-rom/program/pdf/C104/C104-013\\_e.pdf](http://www.jpгу.org/publication/cd-rom/2007cd-rom/program/pdf/C104/C104-013_e.pdf)> (2007).
- Cohen, A. S. & Coe, A. L. New geochemical evidence for the onset of volcanism in the Central Atlantic magmatic province and environmental change at the Triassic–Jurassic boundary. *Geology* **30**, 267–270 (2002).
- Ravizza, G. & Peucker-Ehrenbrink, B. Chemostratigraphic evidence of Deccan volcanism from the marine osmium isotope record. *Science* **302**, 1392–1395 (2003).
- Alvarez, W., Alvarez, L. W., Asaro, F. & Michel, H. V. The end of the Cretaceous: Sharp boundary or gradual transition? *Science* **223**, 1183–1186 (1984).
- Sharma, M., Rosenberg, E. J. & Butterfield, D. A. Search for the proverbial mantle osmium sources to the oceans: Hydrothermal alteration of mid-ocean ridge basalt. *Geochim. Cosmochim. Acta* **71**, 4655–4667 (2007).

26. Lamolda, M. A., Gorostidi, A. & Paul, C. R. C. Quantitative estimates of calcareous nannofossil changes across the Plenus Marls (latest Cenomanian), Dover, England: implications for the generation of the Cenomanian–Turonian Boundary Event. *Cretac. Res.* **15**, 143–164 (1994).
27. Barron, E. J. A. Warm, equable Cretaceous: The nature of the problem. *Earth Sci. Rev.* **19**, 305–338 (1983).
28. Hadas, P. & Mutterlose, J. Calcareous nannofossil assemblages of Oceanic Anoxic Event 2 in the equatorial Atlantic: Evidence of an eutrophication event. *Mar. Micropaleontol.* **66**, 52–69 (2007).
29. Jarvis, I. *et al.* Microfossil assemblages and the Cenomanian–Turonian (late Cretaceous) Oceanic Anoxic Event. *Cretac. Res.* **9**, 3–103 (1988).
30. Selby, D. & Creaser, R. A. Re–Os geochronology of organic rich sediments: An evaluation of organic matter analysis methods. *Chem. Geol.* **200**, 225–240 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Erba for support, D. Tiraboschi for sampling assistance, A. Vogel and K. Jones for help in sample preparation, B. Kendall for discussions and suggestions, and J. Hallows and G. Hatchard for technical assistance. Site 1260B samples were provided by the Integrated Ocean Drilling Program. This study was supported by a Natural Science and Engineering Research Council (NSERC) Discovery Grant. The Radiogenic Isotope Facility at the University of Alberta is supported in part by an NSERC Major Resources Support Grant.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.C.T. ([turgeonsc@ualberta.ca](mailto:turgeonsc@ualberta.ca)).



## METHODS

**Sample preparation.** Large (over ~30 g) samples were powdered with non-metallic methods to avoid potential contamination by osmium-bearing metallic particles. Powdered sediment (0.5 g), representing a 1–2-cm thickness, and a mixed  $^{185}\text{Re}$ – $^{190}\text{Os}$  spike were digested for 48 h at 80 °C in sealed Carius tubes with a  $\text{Cr}^{\text{VI}}\text{O}_3$ – $\text{H}_2\text{SO}_4$  solution. All analyses were performed at the Radiogenic Isotope Facility of the Department of Earth and Atmospheric Sciences, University of Alberta, and are modified from ref. 30.

**Osmium separation.** After digestion, the osmium was separated and purified by solvent extraction three times with 3.5 ml of chloroform added to 3 ml of 9 M HBr and heated overnight to 80 °C. The chloroform was then removed and the HBr was evaporated at 80 °C. The residue was microdistilled with 30  $\mu\text{l}$  of Cr solution (0.2 g of  $\text{CrO}_3$  per ml of 3 M  $\text{H}_2\text{SO}_4$ ) collected in 20  $\mu\text{l}$  of 9 M HBr inside a conical Savillex vial and heated to 80 °C for 3 h. The HBr is then dried and redistilled. Final dry-down of the osmium-bearing HBr was performed at 50 °C under a nitrogen atmosphere.

**Rhenium chromatography.** A 1-ml aliquot of the rhenium-bearing  $\text{CrO}_3$ – $\text{H}_2\text{SO}_4$  solution was diluted with 1 ml of 18.2 M $\Omega$  cm water and the  $\text{Cr}^{\text{VI}}$  reduced to  $\text{Cr}^{\text{III}}$  by bubbling for ~15–30 s with  $\text{SO}_2$  cleaned through water in a gas-washing bottle. Anion chromatography was then performed on disposable polyethylene columns with Eichrom 1-X8 100–200-mesh anion resin. The sample solution was loaded on the column and washed with 4 ml of 1.0 M HCl, 1 ml of 0.2 M  $\text{HNO}_3$  and 1.5 ml of 6 M  $\text{HNO}_3$ , and rhenium was eluted with 4 ml of 6 M  $\text{HNO}_3$ , collected in a clean beaker and evaporated to dryness. The residue was dissolved in 200  $\mu\text{l}$  of 0.05 M  $\text{HNO}_3$ , and further purified with a single DOWEX AG1-X8 <20-mesh anion bead. The 0.05 M  $\text{HNO}_3$  was discarded and replaced with 1 ml of 6 M  $\text{HNO}_3$  overnight, and evaporated to dryness at 80 °C.

**Mass spectrometry.** Purified analytes were loaded onto nickel (for rhenium) and platinum (for osmium) filaments, and isotopic compositions were measured with isotope dilution–thermal ionization mass spectrometry on a VG Sector 54 instrument in negative-ion mode. Isotopic ratios were corrected for instrumental mass fractionation by using  $^{185}\text{Re}/^{187}\text{Re} = 0.59738$  and  $^{192}\text{Os}/^{188}\text{Os} = 3.08261$ , isobaric oxygen interferences, spike and blank contributions. Total procedural blanks were <15 and <0.5 pg for rhenium and osmium, respectively, with a  $^{187}\text{Os}/^{188}\text{Os}$  blank ratio of ~0.20. In-house standard solutions were analysed repeatedly to monitor long-term mass spectrometer reproducibility. Errors for isotope ratios were determined by numerical error propagation and include uncertainties in spike calibration, weighing, and analytical/instrumental measurements, as well as rhenium bias, oxygen and blank corrections.

**Initial seawater osmium isotope calculations.** Initial seawater osmium isotopes were determined from

$$(^{187}\text{Os}/^{188}\text{Os})_{\text{initial}} = (^{187}\text{Os}/^{188}\text{Os})_{\text{measured}} - (^{187}\text{Re}/^{188}\text{Os})_{\text{measured}} \times (e^{\lambda t} - 1)$$

where  $\lambda$  is the  $^{187}\text{Re}$  decay constant ( $\lambda^{187}\text{Re}$ ) of  $1.666 \times 10^{-11} \text{ year}^{-1}$  and  $t$  is the age of the sedimentary rock (93.5 Myr). No adjustments were made for the duration of the OAE2; even with a higher estimate of 0.55 Myr in duration (ref. 10), the initial osmium isotopes would vary by <0.4% on average.

**Mass balance calculations.** The relative mass balance equation used to estimate the contribution of mantle-derived osmium to seawater is  $R_{\text{sw}} = xR_{\text{mtl}} + yR_{\text{chon}} + zR_{\text{cont}}$ , where  $R$  denotes  $^{187}\text{Os}/^{188}\text{Os}$  ratios and the subscript sw represents the calculated initial seawater value, and mtl, chon and cont denote mantle, chondritic and continental sources, respectively. The factors  $x$ ,  $y$  and  $z$  are the relative contributions of each source. Extraterrestrial contributions are assumed here to be negligible<sup>19</sup> and constant for this period, and therefore the relative contribution of mantle-derived osmium can then be estimated by using  $x = (R_{\text{cont}} - R_{\text{sw}})/(R_{\text{cont}} - R_{\text{mtl}})$ . Values for  $R_{\text{cont}}$  and  $R_{\text{mtl}}$  are 1.4 and 0.127, respectively<sup>18</sup>.

# A unifying framework for dinitrogen fixation in the terrestrial biosphere

Benjamin Z. Houlton<sup>1,2,†</sup>, Ying-Ping Wang<sup>3</sup>, Peter M. Vitousek<sup>1</sup> & Christopher B. Field<sup>2</sup>

Dinitrogen (N<sub>2</sub>) fixation is widely recognized as an important process in controlling ecosystem responses to global environmental change, both today<sup>1</sup> and in the past<sup>2</sup>; however, significant discrepancies exist between theory and observations of patterns of N<sub>2</sub> fixation across major sectors of the land biosphere. A question remains as to why symbiotic N<sub>2</sub>-fixing plants are more abundant in vast areas of the tropics than in many of the mature forests that seem to be nitrogen-limited in the temperate and boreal zones<sup>3</sup>. Here we present a unifying framework for terrestrial N<sub>2</sub> fixation that can explain the geographic occurrence of N<sub>2</sub> fixers across diverse biomes and at the global scale. By examining trade-offs inherent in plant carbon, nitrogen and phosphorus capture, we find a clear advantage to symbiotic N<sub>2</sub> fixers in phosphorus-limited tropical savannas and lowland tropical forests. The ability of N<sub>2</sub> fixers to invest nitrogen into phosphorus acquisition seems vital to sustained N<sub>2</sub> fixation in phosphorus-limited tropical ecosystems. In contrast, modern-day temperatures seem to constrain N<sub>2</sub> fixation rates and N<sub>2</sub>-fixing species from mature forests in the high latitudes. We propose that an analysis that couples biogeochemical cycling and biophysical mechanisms is sufficient to explain the principal geographical patterns of symbiotic N<sub>2</sub> fixation on land, thus providing a basis for predicting the response of nutrient-limited ecosystems to climate change and increasing atmospheric CO<sub>2</sub>.

The geographic occurrence of N<sub>2</sub>-fixing organisms in the open ocean<sup>4,5</sup> and in lakes<sup>6</sup> makes sense. Where nitrogen (N) is in low supply, N<sub>2</sub> fixers have an advantage: they can fix N<sub>2</sub> into biomass and thus grow faster than their competitors. In contrast, where N is abundant, N<sub>2</sub> fixation is energetically costly and N<sub>2</sub> fixers are competitively excluded by non-fixing species. N<sub>2</sub>-fixing organisms thereby bring the oceanic inventory of N into equilibrium with N losses over millennia<sup>7</sup>, stabilizing nutrient demands of and supplies to marine primary producers<sup>4</sup>. However, this paradigm is inadequate to predict the global distribution of N<sub>2</sub> fixation in terrestrial environments<sup>8</sup>. Although models based on energetics, nutrients and other factors provide reasonable explanations for the distribution of symbiotic N<sub>2</sub>-fixing plants in succession and as a function of soil fertility within regions, they do not explain the global distribution of N<sub>2</sub> fixation in forests. N<sub>2</sub>-fixing trees (which are predominantly members of the Fabaceae family) are scarce in mature high-latitude forests despite the prevalence of N limitation there<sup>8</sup>; conversely, they account for a significant fraction of communities of many lowland tropical forests<sup>9</sup>, despite the overall N-rich conditions of such ecosystems<sup>10</sup>. Here we build on earlier analyses to resolve the global pattern in terrestrial N<sub>2</sub> fixation, through a combination of empirical data synthesis and economic cost-benefit modelling.

We have developed two new hypotheses for understanding the distribution of N<sub>2</sub> fixation across global ecosystems. The first is that

temperature constrains the distribution of N<sub>2</sub> fixation, contributing to the lack of symbiotic N<sub>2</sub>-fixing trees in mature forests at high latitudes. N<sub>2</sub> fixation is enzymatic and incurs a substantial carbon (C) cost<sup>11</sup>; hence, the rate of N<sub>2</sub> fixation should increase with increasing temperature to some maximum rate. Whether N<sub>2</sub> fixation rates observed for different organisms and environments converge on a similar temperature maximum will be affected by the degree of acclimation to local climatic conditions. Any temperature constraint could reinforce the well-recognized influence of the energy cost of N<sub>2</sub> fixation<sup>12</sup>; the short growing season, relatively low net primary productivity (NPP), and low light availability in mature high-latitude forests could offset the advantages of N<sub>2</sub> fixation over plant N uptake from the soil<sup>13</sup>.

The second hypothesis is that symbiotic N<sub>2</sub>-fixing plants hold an advantage in phosphorus (P) acquisition, as recently described for N<sub>2</sub>-fixing organisms in the P-deficient gyres of the North Atlantic Ocean<sup>14</sup>. N<sub>2</sub> fixers could rely on extracellular phosphatases, a constitutively N-rich (~15% N, ref. 15) class of enzymes involved in the breakdown of organic P (ref. 16), to enhance local P supplies. Once secreted into the soil by plants or microbes, extracellular phosphatases hydrolyze phosphodiester-bonded P (which accounts for 20% to 80% of soil organic P, ref. 17) to phosphate ions that are available for uptake by plant roots<sup>16</sup>. Although many factors control organisms' enzyme activities, phosphatase production increases substantially in response to added N (refs 15, 18), owing to the large N requirement associated with this P acquisition strategy. This hypothesis addresses the second part of the problem, offering an explanation for the persistence of N<sub>2</sub>-fixing plants in mature lowland tropical forests and savannas, many of which seem substantially limited by P (ref. 19).

We found empirical support for both of these hypotheses. Our synthesis of the relationship between temperature and nitrogenase activity, spanning diverse species, strains, latitudes and environments, demonstrated a strong convergent effect of temperature on biochemical N<sub>2</sub> fixation (Fig. 1). Nitrogenase activity (that is, N<sub>2</sub> fixation) reaches a maximum at ~26 °C and decreases at higher temperatures, probably in response to depletion of C supplies. The data conform reasonably well to a single curve, suggesting little evidence for local acclimation across factors. The compiled data yield a slope (activation energy ~103 kJ mol<sup>-1</sup>) that is substantially less than that for the nitrogenase enzyme itself (activation energy ~210 kJ mol<sup>-1</sup>, ref. 20), although substantially greater than the temperature sensitivity of photosynthesis. This suggests that an interaction between energy supply (by means of photosynthesis) and the potential rate of N<sub>2</sub> fixation underlies the observed pattern. Overall, this shows a strong temperature constraint to maximal N<sub>2</sub> fixation rates under cooler climates, reinforcing the high energetic cost of fixing N<sub>2</sub>, and reducing the marginal return of N per unit of C

<sup>1</sup>Biological Sciences, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Department of Global Ecology, Carnegie Institution of Washington, Stanford, California 94305, USA.

<sup>3</sup>CSIRO Marine and Atmospheric Research and Centre for Australian Weather and Climate Research, Aspendale VIC 3195, Victoria, Australia. <sup>†</sup>Present address: Department of Land, Air and Water Resources, University of California, Davis, California 95616, USA.

investment by plants in  $N_2$  fixation compared to soil N uptake. In other words, as temperature declines, more enzyme is needed to achieve a given rate of  $N_2$  fixation.

In support of the second hypothesis, phosphatase production rates were three times higher in soils sampled from beneath plants with known capacity to fix  $N_2$  compared to that beneath non-fixing species (Fig. 2;  $P < 0.001$ ;  $n = 25$ ). This synthesis includes data from known rhizobial and actinorhizal  $N_2$ -fixing plants, differing biogeochemical conditions and plant functional growth forms (that is, herbaceous and woody vegetation), and across temperate/tropical latitudes. The higher phosphatase concentrations beneath these putative  $N_2$  fixers probably function locally to mobilize organic P forms to organisms in low P environments. The origin of the phosphatase is uncertain; it could be produced directly by  $N_2$ -fixer roots<sup>21</sup>, by microbes feeding on the N-rich litter of  $N_2$  fixers, or (most probably) by both. In any case, the addition of N stimulates phosphatase production<sup>15,18</sup>, thereby stimulating the availability of P and indicating an N-rich strategy of P acquisition that is particularly suited to  $N_2$ -fixing species.

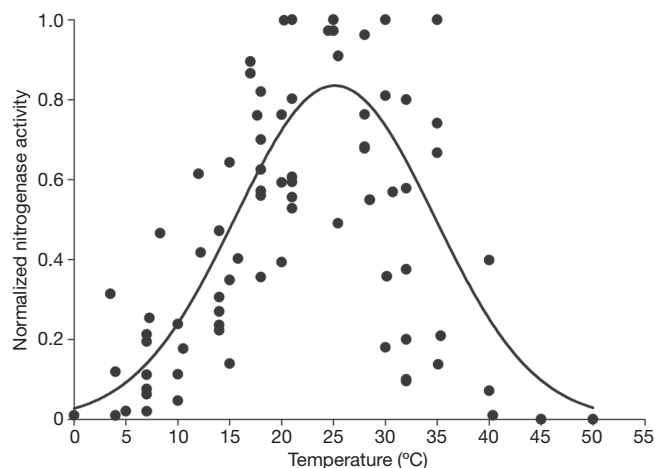
We therefore propose that temperature amplifies the energetic constraint to  $N_2$ -fixing plants in mature high-latitude forests where N is often limiting, and also that  $N_2$  fixers are adept at acquiring P by means of phosphatase enzymes, thus providing a means for their persistence in P-limited tropical ecosystems. Taken together and when combined with other costs of fixing  $N_2$  identified previously<sup>11–13</sup> (for example, light, demand for other nutrients, and herbivory), we suggest that these mechanisms can explain patterns of  $N_2$ -fixing plants and  $N_2$ -fixation fluxes on land.

We addressed the sufficiency of this proposed framework for resolving the pattern of  $N_2$  fixers across diverse terrestrial environments and conditions using a terrestrial biogeochemical model<sup>22</sup> (Supplementary Information) that simulates the economics (costs and benefits) of C, N and P acquisition and competition between  $N_2$  fixers and non-fixers according to the resource optimization paradigm<sup>23</sup>. The model consists of 8 C pools, 9 N pools and 12 P pools, 6 of each of which are divided equally into the plants ( $N_2$  fixer and non-fixer) and the remainder are in the soil. To calibrate the model, we varied maximal C, N and P uptake rates until all pool sizes at equilibrium agreed with empirical observations (average conditions) for four diverse terrestrial biomes: boreal forest, temperate forest, lowland tropical forest and tropical savanna (see Supplementary Table 1). The model is driven by the observed monthly soil mean

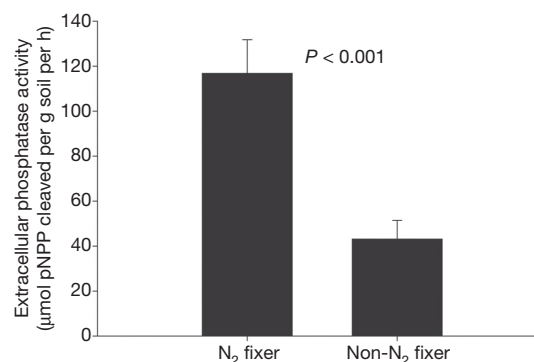
temperature and maximal NPP of each biome, which implicitly depends on available light and water. Competitive outcomes are determined by NPP gains per unit of resource investment. We assume that NPP varies as a function of light interception above-ground and nutrient acquisition below-ground; the model predicts optimal strategies for maximizing NPP, adjusting N acquisition ( $N_2$  fixation versus soil N uptake), P acquisition (phosphatase investment versus uptake of microbially mineralized P) and C allocation (shoots versus roots).

Building on our base model<sup>22</sup>, we included the constraint of temperature on maximal  $N_2$  fixation rates (Fig. 1; see also Methods equation (4)) and two different strategies of P acquisition by means of phosphatase investments: first, a global commons model in which P mineralized by phosphatase enzymes enters a common pool that is available to all plant and microbial competitors equally; and second, an individual-based strategy in which the plant producing phosphatase gets 'first crack' at the mineralized P (ref. 24). We explicitly assume a fixed cost of 15 g N per g P (ref. 15), the required investment of N in phosphatase production by roots or the microbial community. We implicitly assume that the N invested by the microbial community in phosphatase production is supplied by plant litter. We ran the model for three different situations to test our framework: (1) temperature-dependent  $N_2$  fixation (on the basis of the empirical data in Fig. 1), individual based; (2) temperature-dependent  $N_2$  fixation, global commons; and (3) constant  $N_2$  fixation, individual based.

This model-based analysis agrees quantitatively with our proposed framework (simulation (1), described above) for controls of terrestrial  $N_2$  fixation globally. The inclusion of temperature in our model results in the exclusion of  $N_2$  fixers from mature forests at high latitudes (simulations (1) and (2), Fig. 3a, b). Despite profound N limitation (Fig. 3i, j), temperature limits  $N_2$  fixation rates in temperate and boreal forests; here, the investment of C in soil N uptake yields the greatest NPP return. A temperature increase of 10 °C can alleviate this constraint on  $N_2$  fixers, suggesting a potential interaction between climate warming and  $N_2$  fixation at higher latitudes. Furthermore, according to our model, higher rates of N loss leading to more profound N limitation can overcome the energetic limitations on  $N_2$  fixation in temperate forests (Supplementary Fig. A2). This latter result is consistent with the transient presence of  $N_2$ -fixing plants in early succession and disturbed temperate ecosystems<sup>8</sup> that have lost substantial quantities of N (refs 25, 26). Thus, our framework is able to reconcile both the presence and the absence of  $N_2$ -fixing plants in extra-tropical forests, and suggests that  $N_2$  fixation



**Figure 1 | Temperature dependence of terrestrial nitrogenase activity.** We compiled these data from the primary literature and normalized them to the maximal activity observed for each individual study (see Supplementary Information). The data set spans a wide range of terrestrial conditions, latitudes, strains and species of  $N_2$ -fixing organisms, but can be fitted with an empirical equation with an optimal temperature of 25.2 °C ( $n = 94$ ,  $r^2 = 0.55$ ).



**Figure 2 | Phosphatase enzyme rates in soils with and without  $N_2$ -fixing plants.** We compiled these data from the primary literature, which included soils collected from actinorhizal and leguminous  $N_2$ -fixing plant species, herbaceous and woody plants, and across temperate to tropical latitudes (see Supplementary Information). Extracellular phosphatase fluxes were significantly higher in the presence of  $N_2$ -fixing plants (two-tailed  $t$ -test,  $P < 0.001$ ;  $n = 25$ ) compared to non-fixers only. pNPP, *para*-Nitrophenyl Phosphate. Error bars represent s.e.m.



can alleviate N limitation (for example, simulation 3; Fig. 3i, j) where it does occur.

Phosphatase enzymes and associated C, N and P interactions seem to be central to resolving the relatively high abundance of N<sub>2</sub>-fixing plants in the tropics (Fig. 3c, d). Despite profound P limitation, our model simulates large contributions of N<sub>2</sub>-fixing plants (from 20% to 50% of NPP) to tropical forest and savanna ecosystems when the P released by phosphatase is captured by the plant that produced the phosphatase (see cluster roots<sup>24</sup>; simulations (1) and (3); Fig. 3c, d). Our simulations indicate that newly fixed N is vital to stimulating phosphatase production and organic P mineralization rates; non-fixing plants and free-living microbes must rely on existing soil N pools, thus limiting their capacity for investment in N-rich phosphatase enzymes. According to our model, this proposed benefit to N<sub>2</sub>-fixing plants sufficiently offsets the integrated costs (for example, C and light) associated with fixing N<sub>2</sub> in the tropics, where monthly mean temperatures are close to the optimal rate of N<sub>2</sub> fixation (Fig. 1) and P is often in low supply. If P is shared among competitors equally (that is, plants, microbes and geochemical sinks; global commons), this advantage is lost (simulation (2)), further illustrating the importance of below-ground competition for P in the tropics.

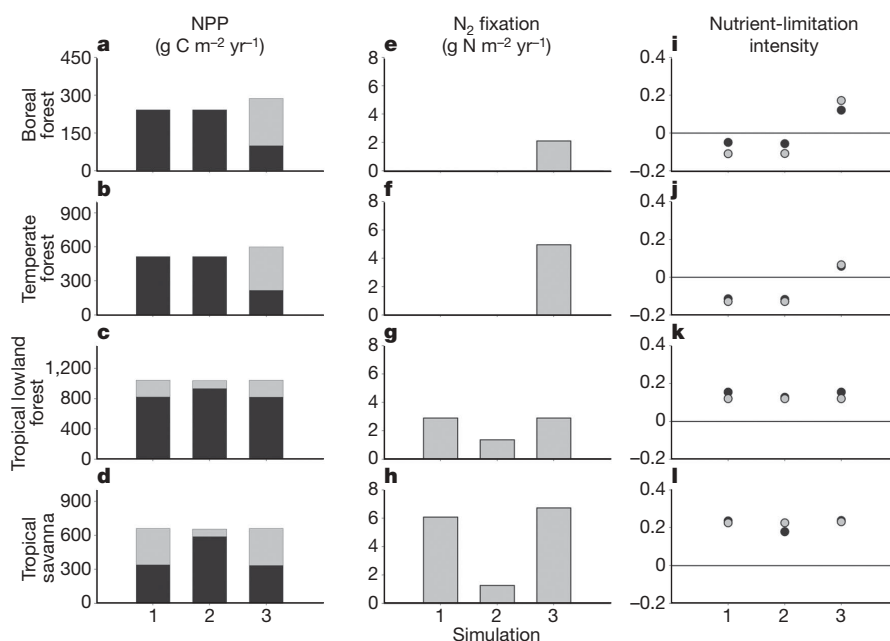
As a whole, our biogeochemical and biophysical mechanisms offer a unified explanation for global-scale patterns of terrestrial N<sub>2</sub> fixation; this working framework (which includes interactions between energy (C, light), nutrients (N, P) and temperature) seems sufficient to resolve the distribution of N<sub>2</sub>-fixing plants across major forests worldwide. In tropical lowland forests, our model suggests that competition for P is central to the persistence of N<sub>2</sub> fixers. Future work could emphasize the sources and fates of P in different tropical soils, differences among the P-acquisition strategies of organisms, plant versus microbial phosphatase production rates, and the extent to which phosphatase producers receive a benefit over their competitors. Outside the tropics, temperature seems to impose a principal constraint to N<sub>2</sub> fixers in mature forests, whereas energy, other resources or plant traits may be important in controlling the abundance of N<sub>2</sub> fixers in other terrestrial high-latitude biomes, such as prairies and grasslands.

Finally, our proposed global framework (described by simulation (1)) and corresponding model indicate substantial symbiotic N<sub>2</sub> fixation rates by trees in lowland tropical forests and savannas ( $\sim 20$  to  $60 \text{ kg N ha}^{-1} \text{ yr}^{-1}$ ) and little to none in mature temperate and boreal forests (Fig. 3e–h). Both the magnitude and pattern of these modelled N inputs agree qualitatively with the relative abundance of potentially N<sub>2</sub>-fixing legumes in those ecosystems<sup>3,9</sup>. They also agree with estimates of N<sub>2</sub> fixation rates extrapolated by way of environmental proxy data to continental scales<sup>27</sup>, and the ‘missing N’ inputs observed for lowland tropical forest watersheds (see, among others, ref. 28 for example). We have not addressed free-living N<sub>2</sub> fixation here; future work might consider this pathway in the context of our framework and its potential magnitude globally. Furthermore, in addition to our framework, factors such as herbivory, trace metal availability, life history traits<sup>9</sup> and succession may help explain observations<sup>29</sup> for spatial and temporal patchiness in nodulation and in the contribution of N<sub>2</sub> fixation to the N economy of an individual tree<sup>30</sup> in tropical ecosystems.

## METHODS SUMMARY

We performed boolean keyword searches using the ISI Web of Science to gain information on the temperature dependence of nitrogenase activity and phosphatase enzyme rates in soils sampled beneath N<sub>2</sub> fixers and non-fixers. For temperature effects on nitrogenase, we used the keywords ‘temperature’ and ‘nitrogenase’; for phosphatase activities we used the keywords ‘phosphatase’, ‘nitrogenase’, ‘nitrogen’ and ‘fix\*’. Data for the effects of temperature on nitrogenase activity were normalized to the maximum rate observed for each individual experiment. Phosphatase data were lumped by functional group (that is, N<sub>2</sub> fixer versus non-N<sub>2</sub> fixer) and analysed for statistical significance using a Student’s *t*-test. Details of the outcome of our literature search can be found in Methods.

Our model simulates the competition between N<sub>2</sub> fixers and non-fixers for light above-ground and nutrients below-ground. Details of the model equations can be found in the Methods and have been described previously<sup>22</sup>; details of the model calibration and simulation schemes are in the Supplementary Information. With the exception of fixing atmospheric N<sub>2</sub>, model parameters are identical for N<sub>2</sub> fixers and non-fixers. Both N<sub>2</sub> fixers and non-fixers can take up soil N and can use any acquired N to produce phosphatases to increase soil labile-P availability. The successful invasion and establishment of N<sub>2</sub> fixers into a non-fixer dominant ecosystem depends on the return in NPP of its C investment into N<sub>2</sub> fixation



**Figure 3 | Model results for different hypotheses across terrestrial biomes at steady state.** a–l, The hypotheses tested are: (1), temperature-dependent N<sub>2</sub> fixation and individual-based P acquisition strategy (consistent with our proposed framework); (2), temperature-dependent N<sub>2</sub> fixation and global-commons P acquisition strategy; and (3), constant N<sub>2</sub> fixation (no temperature effect) and individual-based P acquisition strategy. a–d, NPP by

N<sub>2</sub> fixers (grey) and non-fixers (black); e–h, the N<sub>2</sub> fixation rate; i–l, the nutrient-limitation intensity for N<sub>2</sub> fixers (grey) and non-fixers (black). Nutrient-limitation intensity was calculated as the difference in N availability minus P availability, such that positive values indicate P limitation and negative values indicate N limitation.

when NPP is N-limiting and also on the benefit of investing N into phosphatase production to reduce P limitation and increase NPP when P is limiting. All pool sizes of C, N or P at steady state are proportional to the uptake rate of C, N or P, respectively. The model has two unique and one non-unique equilibrium solutions for the parameters we used in this study.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 5 December 2007; accepted 24 April 2008.

Published online 18 June 2008.

- Hungate, B. A., Dukes, J. S., Shaw, M. R., Luo, Y. & Field, C. B. Nitrogen and climate change. *Science* **302**, 1512–1513 (2003).
- Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature* **387**, 272–275 (1997).
- Crews, T. E. The presence of nitrogen fixing legumes in terrestrial communities: Evolutionary vs ecological considerations. *Biogeochemistry* **46**, 233–246 (1999).
- Redfield, A. C. The biological control of chemical factors in the environment. *Am. Sci.* **46**, 205–221 (1958).
- Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. & Dunne, J. P. Spatial coupling of nitrogen inputs and losses in the ocean. *Nature* **445**, 163–167 (2007).
- Schindler, D. W. Eutrophication and recovery in experimental lakes: implications for lake management. *Science* **184**, 897–899 (1974).
- Deutsch, C., Sigman, D. M., Thunell, R. C., Meckler, A. N. & Haug, G. H. Isotopic constraints on glacial/interglacial changes in the oceanic nitrogen budget. *Glob. Biogeochem. Cycles* **18**, GB4012 (2004).
- Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry* **13**, 87–115 (1991).
- ter Steege, H. et al. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* **443**, 444–447 (2006).
- Martinelli, L. A. et al. Nitrogen stable isotopic composition of leaves and soil: Tropical versus temperate forests. *Biogeochemistry* **46**, 45–65 (1999).
- Gutschick, V. P. Evolved strategies in nitrogen acquisition by plants. *Am. Nat.* **118**, 607–637 (1981).
- Vitousek, P. M. & Field, C. B. Ecosystem constraints to symbiotic nitrogen fixers: a simple model and its implications. *Biogeochemistry* **46**, 179–202 (1999).
- Rastetter, E. B. et al. Resource optimization and symbiotic nitrogen fixation. *Ecosystems* **4**, 369–388 (2001).
- Dyhrman, S. T. et al. Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* **439**, 68–71 (2006).
- Treseder, K. K. & Vitousek, P. M. Effects of soil nutrient availability on investment in acquisition of N and P in Hawaiian rain forests. *Ecology* **82**, 946–954 (2001).
- Duff, S. M. G., Sarath, G. & Plaxton, W. C. The role of acid phosphatases in plant phosphorus metabolism. *Physiol. Plant.* **90**, 791–800 (1994).
- McGill, W. B. & Cole, C. V. Comparative aspects of cycling of organic C, N, S and P through soil organic matter. *Geoderma* **26**, 267–286 (1981).
- Olander, L. P. & Vitousek, P. M. Regulation of soil phosphatase and chitinase activity by N and P availability. *Biogeochemistry* **49**, 175–190 (2000).
- Vitousek, P. M. & Sanford, R. L. Jr. Nutrient cycling in moist tropical forest. *Annu. Rev. Ecol. Syst.* **17**, 137–167 (1986).
- Ceuterick, F. et al. Effect of high pressure, detergents and phospholipase on the break in the Arrhenius plot of *Azotobacter* nitrogenase. *Eur. J. Biochem.* **87**, 401–407 (1978).
- Kamh, M., Abdou, M., Chude, V., Wiesler, F. & Horst, W. J. Mobilization of phosphorus contributes to positive rotational effects of leguminous cover crops on maize grown on soils from northern Nigeria. *J. Plant Nutr. Soil Sci.* **165**, 566–572 (2002).
- Wang, Y. P., Houlton, B. Z. & Field, C. B. A model of biogeochemical cycles of carbon, nitrogen, and phosphorus including symbiotic nitrogen fixation and phosphatase production. *Glob. Biogeochem. Cycles* **21**, GB1018 (2007).
- Bloom, A. J., Chapin, F. S. III & Mooney, H. A. Resource limitation in plants—an economic analogy. *Annu. Rev. Ecol. Syst.* **16**, 363–392 (1985).
- Lambers, H., Shane, M. W., Cramer, M. D., Pearce, S. J. & Veneklaas, E. J. Root structure and functioning for efficient acquisition of phosphorus: matching morphological and physiological traits. *Ann. Bot.* **98**, 693–713 (2006).
- Bormann, F. H. & Likens, G. E. *Pattern and Process in a Forested Ecosystem* (Springer, Berlin, 1979).
- Houlton, B. Z. et al. Nitrogen dynamics in ice storm-damaged forest ecosystems: implications for nitrogen limitation theory. *Ecosystems* **6**, 431–443 (2004).
- Cleveland, C. C. et al. Global patterns of terrestrial biological nitrogen (N<sub>2</sub>) fixation in natural ecosystems. *Glob. Biogeochem. Cycles* **13**, 623–645 (1999).
- Chestnut, T. J., Zarin, D. J., McDowell, W. H. & Keller, M. A nitrogen budget for late-successional hillslope tabonuco forest, Puerto Rico. *Biogeochemistry* **46**, 85–108 (1999).
- Pons, T. L., Perreijn, K., van Kessel, C. & Werger, M. J. A. Symbiotic nitrogen fixation in a tropical rainforest: <sup>15</sup>N natural abundance measurements supported by experimental isotopic enrichment. *New Phytol.* **173**, 154–167 (2007).
- Sprent, J. I. & Raven, J. A. Evolution of nitrogen-fixing symbioses. *Proc. Royal Society Edinburgh Section B-Biological Sciences* **85**, 215–237 (1985).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by the National Science Foundation, CSIRO, the Australian Greenhouse Office, the David and Lucile Packard Foundation, and the US Department of Energy.

**Author Contributions** B.Z.H. wrote the initial manuscript. Y.P.W. and B.Z.H. performed the model simulations. All authors discussed the approach, organization and results, and developed and improved the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.Z.H. ([bzhoulton@ucdavis.edu](mailto:bzhoulton@ucdavis.edu)).

## METHODS

**Data compilation.** Using our keyword search (see Methods Summary), we identified six studies<sup>31–36</sup> in the primary literature that examined nitrogenase activity across a range of temperatures in the laboratory, spanning multiple species, strains and growth forms of  $N_2$  fixers, and terrestrial latitudes ranging from the extra-tropics to the tropics. We identified five studies<sup>37–41</sup> in the primary literature that evaluated soil phosphatase amounts in a paired design between known  $N_2$  fixers and non-fixers, including data from temperate and tropical biomes, different biogeochemical conditions and across functional growth forms ranging from annual forbs to evergreen trees.

**Curve-fitting statistics.** We fitted a three-parameter equation of  $y_{N,fix} = \exp[a + bT_s(1 - 0.5T_s/c)]$  to all data points in Fig. 1 using the nonlinear regression routine in Sigmaplot (version 10.0, 2006 Systat Software Inc.), where  $T_s$  is the rooting zone temperature. The best estimates of three parameters and one s.e.m. are  $a = -3.62 \pm 0.52$ ,  $b = 0.27 \pm 0.04$  and  $c = 25.15 \pm 0.66$  ( $n = 94$ ,  $r^2 = 0.55$ ). The fitted equation was then normalized to give a value of  $y_{N,fix} = 1.0$  at  $T_s = c$  by multiplying by a scaling factor of 1.25.

**The model.** Our model simulates the competition between  $N_2$  fixers and non-fixers for light above-ground and nutrients below-ground. Details of the model can be found in ref. 22. Changes in C, N or P in each pool are calculated by integrating a set of differential equations numerically. The two unique equilibria are: exclusion of the  $N_2$  fixer, with its NPP decreasing to zero (see Fig. 3a, simulations (1) and (2)); or invasion and establishment in the community (see Fig. 3a, simulation (3)). In the latter case,  $N_2$  fixers will coexist with non-fixers. In case of a non-unique equilibrium solution, the initial fractions of  $N_2$  fixers and its share of total NPP will be maintained (see Fig. 3c, d, simulation (2)).

**Competition for light.** We assumed that the NPP of  $N_2$  fixers and non-fixers is proportional to the intercepted photosynthetically active radiation by its leaves, and that all leaves (either  $N_2$  fixers or non-fixers) are randomly dispersed within the canopy. Therefore, the amount of intercepted light is proportional to the total leaf area of each species. NPP of non- $N_2$  fixers ( $F_{C,nf}$ ) or  $N_2$  fixers ( $F_{C,fix}$ ) is modelled as:

$$F_{C,i} = U_{Cmax} \phi(x_{N,i}, x_{P,i}) \frac{C_{i,1}}{\sum_i C_{i,1}} \left( 1 - \exp(-k_1 \sigma_1 \sum_i C_{i,1}) \right) \quad (1)$$

where  $U_{Cmax}$  is a seasonally varying maximum NPP that reflects the seasonal variation of availability of light and water,  $k_1$  is the extinction coefficient ( $= 0.5$ ),  $\sigma_1$  is the specific leaf area ( $m^2(g C)^{-1}$ ) and  $i = 'fix'$  for  $N_2$  fixers or  $'nf'$  for non-fixers. Function  $\phi(x_{N,i}, x_{P,i})$  describes the dependence of NPP on N availability ( $x_{N,i}$ ) or P availability ( $x_{P,i}$ ). Nutrient availability is calculated as:

$$x_{N,i} = \frac{n_{i,1}}{n_{i,1} + K_n}$$

$$x_{P,i} = \frac{p_{i,1}}{p_{i,1} + K_p}$$

where  $K_n$  and  $K_p$  are two empirical parameters,  $n_{i,1}$  and  $p_{i,1}$  are the leaf N/C and P/C ratios of plant  $i$ , respectively. Nutrient limitation intensity (see Fig. 3), calculated as  $x_{N,i} - x_{P,i}$ , is positive when P is limiting and is negative when N is limiting.

**Competition for soil mineral nitrogen.** We assumed that all roots have equal access to soil mineral N and that the rate of root N uptake by  $N_2$  fixers or non-fixers is proportional to their total root length in the soil. When the integrated C cost of root N uptake exceeds the C cost of symbiotic  $N_2$  fixation, a fraction of the NPP allocated to root growth by  $N_2$  fixers will be used to fix atmospheric  $N_2$ .

As described previously<sup>13</sup>, we calculated the C cost of N uptake ( $\lambda_{Nup}$ ) as the increase in NPP by plants for an extra amount of C invested to grow more leaves divided by the increase in NPP if that extra amount of C were allocated to roots. That is,

$$\lambda_{Nup} = \frac{\frac{\partial F_{C,fix}}{\partial C_{i,1}}}{\frac{\partial F_{Nup,fix}}{\partial (F_{Nup,fix} C_{i,fix})}} = f_{Nup} \frac{\frac{\partial F_{C,fix}}{\partial C_{i,1}}}{\frac{\partial F_{Nup,fix}}{\partial C_{i,fix}}} = f_{Nup} \lambda_{Nup}^* \quad (2)$$

where  $F_{C,fix}$  and  $F_{Nup,fix}$  are the net primary production ( $g C m^{-2} yr^{-1}$ ) and N uptake ( $g N m^{-2} yr^{-1}$ ) of  $N_2$  fixers,  $f_{Nup}$  is the fraction of roots of the  $N_2$  fixers that are not nodulated ( $= 1$  for non-fixers), and  $\lambda_{Nup}^*$  is the C cost of N uptake if none of the roots are nodulated ( $g C (g N)^{-1}$ ).  $C_{i,fix}$  and  $C_{r,fix}$  are the amount leaf and roots of  $N_2$  fixers, respectively, in  $g C m^{-2}$ . The fraction of roots that are

nodulated for fixing  $N_2$  ( $f_{N,fix}$ ) is modelled as:

$$f_{N,fix} = 1 - f_{Nup} = 1 - \min \left( 1.0, \frac{\lambda_{N,fix}}{\lambda_{Nup}^*} \right) \quad (3)$$

where  $\lambda_{N,fix}$  is the carbon cost of  $N_2$  fixation ( $= 6.8 g C (g N)^{-1}$ ).

The rate of symbiotic  $N_2$  fixation ( $F_{N,fix}$ ) is then modelled as:

$$F_{N,fix} = y_{N,fix} f_{N,fix} C_{r,fix} \quad (4)$$

Where  $y_{N,fix}$  is the rate of  $N_2$  fixed per unit of nodulated root C ( $g N (g C)^{-1} yr^{-1}$ ) and is a function of soil temperature (see Fig. 1).

**Competition for soil phosphorus.** The rate of labile soil P uptake by roots of  $N_2$  fixers or non-fixers is assumed to be proportional to its total root length. The rate of biochemical P mineralization increases with the difference between the N cost of labile soil P uptake and the N cost of phosphatase production ( $= 15 g N$  per g P). That is,

$$F_{ptase,i} = U_{ptase} (1 - 5 \exp(-0.07 P_o)) \frac{\lambda_{pup,i} - \lambda_{ptase}}{\lambda_{pup,i} - \lambda_{ptase} + K_{ptase}} \quad (5)$$

where  $U_{ptase}$  is the maximum rate of biochemical P mineralization ( $= 0.01 g P m^{-2} yr^{-1}$ ),  $P_o$  is the total amount of organic P in soil ( $g P m^{-2}$ ),  $\lambda_{pup,i}$  is the N cost of uptake of labile P in soil by roots of  $N_2$  fixer ( $i = fix$ ) or non-fixer ( $i = nf$ ), and  $\lambda_{ptase}$  is the N cost of phosphatase production ( $= 15 g N$  per g P). The second term on the right-hand side of the equation accounts for the reduction of biochemical P mineralization rate as a result of decreasing amounts of organic P to be cleaved by phosphatases in the soil. We also assumed that up to 20% of soil organic P can be cleaved. The N cost of phosphorus uptake ( $\lambda_{pup,i}$ ) was estimated as the increase in NPP for a further increase in P uptake divided by the increase in NPP per unit extra N uptake<sup>22</sup>. That is,

$$\lambda_{pup,i} = \frac{\partial F_{C,i} / \partial F_{P,i}}{\partial F_{C,i} / \partial F_{Nup,i}} \quad (6)$$

where  $F_{P,i}$  is the rate of P uptake by  $N_2$  fixers ( $i = fix$ ) or non-fixers ( $i = nf$ ) ( $g P m^{-2} yr^{-1}$ ), and  $F_{C,i}$  and  $F_{Nup,i}$  are NPP ( $g C m^{-2} yr^{-1}$ ) and N uptake rate by roots ( $g N m^{-2} yr^{-1}$ ), respectively. Phosphatase production ( $F_{ptase,i} > 0$ ) will start when  $\lambda_{pup,i} > \lambda_{ptase}$ .

Both  $N_2$  fixer and non-fixers can produce phosphatases. When the biochemically mineralized P is not shared between  $N_2$  fixers and non-fixers, the biochemically mineralized P by  $N_2$  fixers or non-fixers is all taken up by the phosphatase producers; conversely, when the biochemically mineralized P is shared, the biochemically mineralized P enters the labile soil P pool that is assessable to both competitors equally.

- Coxson, D. S. & Kershaw, K. A. Rehydration response of nitrogenase activity and carbon fixation in terrestrial *Nostoc commune* from *Stipa-Bouteloua* grassland. *Can. J. Bot.* **61**, 2658–2668 (1983).
- Chapin, D. M., Bliss, L. C. & Bledsoe, L. J. Environmental regulation of nitrogen fixation in a high arctic lowland ecosystem. *Can. J. Bot.* **69**, 2744–2755 (1991).
- Roper, M. M. Straw decomposition and nitrogenase activity ( $C_2H_2$  reduction): Effects of soil moisture and temperature. *Soil Biol. Biochem.* **17**, 65–71 (1985).
- Chan, Y.-K. Temperature response of an associative  $N_2$ -fixing *Pseudomonas* species in pure culture. *Can. J. Microbiol.* **37**, 715–718 (1991).
- Schomburg, H. H. & Weaver, R. W. Nodulation, nitrogen fixation, and early growth of arrowleaf clover in response to root temperature and starter nitrogen. *Agron. J.* **84**, 1046–1050 (1992).
- Liengen, T. & Olsen, R. A. Seasonal and site-specific variations in nitrogen fixation in a high arctic area, Ny-Alesund, Spitsbergen. *Can. J. Microbiol.* **43**, 759–769 (1997).
- Zou, X. M., Binkley, D. & Caldwell, B. A. Effects of dinitrogen fixing trees on phosphorus biogeochemical cycling in contrasting forests. *Soil Sci. Soc. Am. J.* **59**, 1452–1458 (1995).
- Giardina, C. P., Huffman, S., Binkley, D. & Caldwell, B. A. Alders increase soil phosphorus availability in a Douglas-fir plantation. *Can. J. Forest Res.* **25**, 1652–1657 (1995).
- Allison, S. D., Nielsen, C. & Hughes, R. F. Elevated enzyme activities in soils under the invasive nitrogen-fixing tree *Falcataria moluccana*. *Soil Biol. Biochem.* **38**, 1537–1544 (2006).
- Caldwell, B. A. Effects of invasive scotch broom on soil properties in a Pacific coastal prairie soil. *Appl. Soil Ecol.* **32**, 149–152 (2006).
- Nuruzzaman, M., Lambers, H., Bolland, M. D. A. & Veneklaas, E. J. Distribution of carboxylates and acid phosphatase and depletion of different phosphorus fractions in the rhizosphere of a cereal and three grain legumes. *Plant Soil* **281**, 109–120 (2006).



# Coherent ecological dynamics induced by large-scale disturbance

Timothy H. Keitt<sup>1</sup>

Aggregate community-level response to disturbance is a principle concern in ecology because post-disturbance dynamics are integral to the ability of ecosystems to maintain function in an uncertain world. Community-level responses to disturbance can be arrayed along a spectrum ranging from synchronous oscillations where all species rise and fall together, to compensatory dynamics where total biomass remains relatively constant despite fluctuations in the densities of individual species<sup>1</sup>. An important recent insight is that patterns of synchrony and compensation can vary with the timescale of analysis<sup>2</sup> and that spectral time series methods can enable detection of coherent dynamics that would otherwise be obscured by opposing patterns occurring at different scales<sup>3</sup>. Here I show that application of wavelet analysis to experimentally manipulated plankton communities reveals strong synchrony after disturbance. The result is paradoxical because it is well established that these communities contain both disturbance-sensitive and disturbance-tolerant species leading to compensation within functional groups<sup>4,5</sup>. Theory predicts that compensatory substitution of functionally equivalent species should stabilize ecological communities<sup>6–10</sup>, yet I found at the whole-community level a large increase in seasonal biomass variation. Resolution of the paradox hinges on patterns of seasonality among species. The compensatory shift in community composition after disturbance resulted in a loss of cold-season dominants, which before disturbance had served to stabilize biomass throughout the year. Species dominating the disturbed community peaked coherently during the warm season, explaining the observed synchrony and increase in seasonal biomass variation. These results suggest that theory relating compensatory dynamics to ecological stability needs to consider not only complementarity in species responses to environmental change, but also seasonal complementarity among disturbance-tolerant and disturbance-sensitive species.

I analysed time series data of 10 dominant zooplankton species sampled over a 17-year period from Little Rock Lake (LRL), Wisconsin, United States (LRL zooplankton data, North Temperate Lakes Long-Term Ecological Research programme). Only crustaceans were included; however, these have been shown to dominate zooplankton biomass in this system<sup>5</sup>. The experimental treatment effect was applied as a transient 'press' perturbation: the lake was separated into two basins at the beginning of the study and pH was artificially reduced in one of the basins from 1985 to 1990, after which the basins remained separated and pH was allowed to return to baseline conditions<sup>11</sup>. Laboratory assays indicate variation among LRL zooplankton species in their tolerance to acidification; it is well established that interactions between these species form a complex web of positive and negative effects<sup>4,5</sup>. The consequences of these complex interactions for community-level variation have remained enigmatic as we have previously lacked methods

sufficiently sensitive to isolate clear evidence for whole-community synchrony or compensation across multiple scales.

Using a multivariate wavelet analysis (see Methods), I found a marked contrast in community dynamics between reference and treatment basins (Fig. 1). The reference basin had a weak trend towards synchrony at sub-annual timescales and a trend towards compensation at annual and longer scales. Short bursts of both synchronous and compensatory dynamics were observed within the reference basin. Some caution is warranted, however, in interpreting short runs of significant values as they often represent false positives<sup>12</sup>. In contrast, the treatment basin exhibited a powerful synchronous annual-scale signal throughout the period of pH reduction and over much of the recovery period. Notably, the pattern of synchrony appears at the beginning of the time series before separation of the basins. Although this observation might suggest latent difference between the basins unrelated to disturbance, I believe this conclusion is unwarranted for several reasons. First, there is an intrinsic limitation in how precisely wavelet results can be localized in time<sup>13</sup> and therefore results at the initial boundary necessarily contain information from subsequent years that were subject to the treatment effect. Second, results near the beginning and end of the time series can be influenced by truncation of the wavelet and so should be interpreted with caution. Third, statistically significant differences in modulus ratios between basins only appear much later in the time series data (Fig. 2), indicating that differences between basins during the first several years were slight. During the period when modulus ratios were significantly different between basins, the treatment basin showed marked changes in seasonal biomass dynamics (Fig. 2). Wavelet analysis of biomass summed across species shows that synchrony in the treatment basin resulted in a greater than fourfold increase in total annual-scale biomass variation. The increase in biomass variation was also apparent in the raw untransformed time series data: the median absolute deviations of total biomass in the reference and treatment basins were 77.4 and 106.9  $\mu\text{g l}^{-1}$ , respectively. Increased biomass variation was furthermore associated with a nearly two-month phase shift in the timing of annual peak biomass between the two basins.

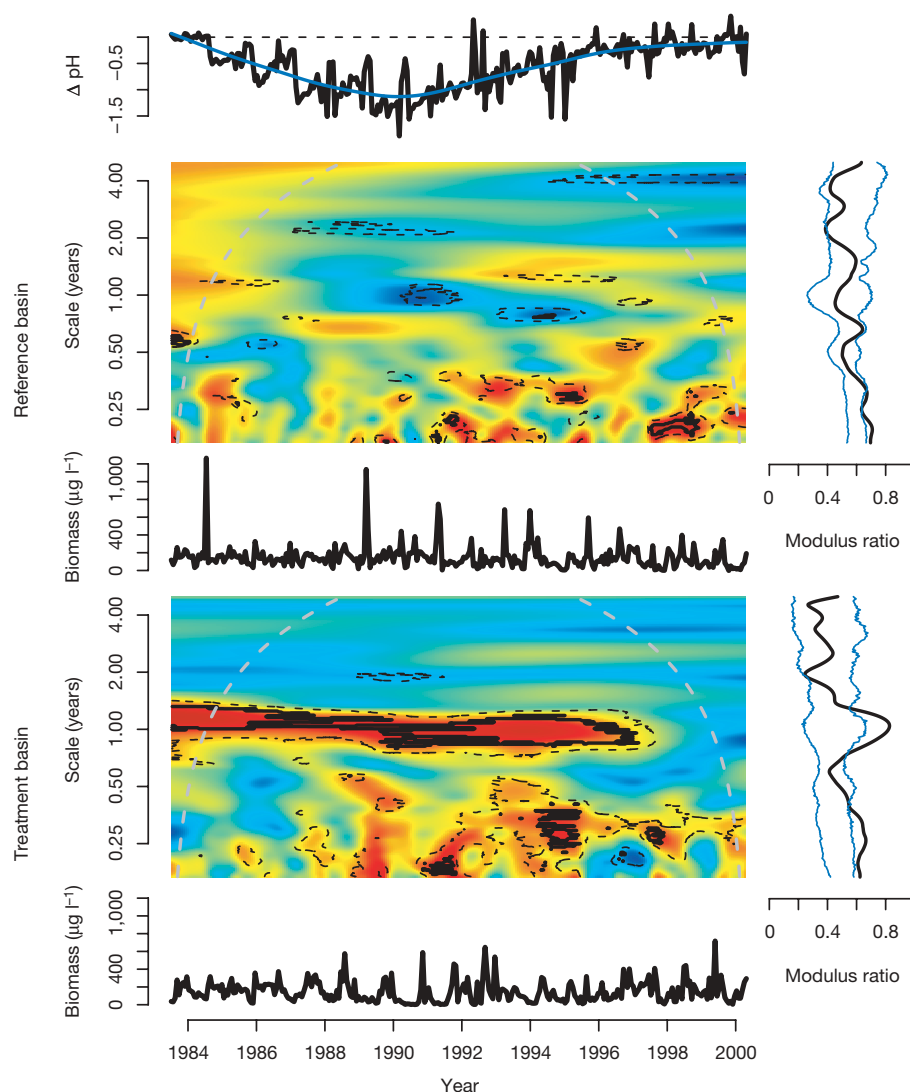
The strong phase shift in total biomass suggests alteration of species phase relationships (Fig. 3 and Supplementary Fig. 1). Individual phase relationships can be examined by plotting the wavelet coefficients as vectors in the complex plane (Fig. 3). Vector pairs pointing in the same direction indicate perfect synchrony between species; those in the opposite direction, perfect compensation. Vector lengths (here log-transformed to aid visualization) scale with the difference between minimum and maximum biomass during the year. During the period of maximal difference in modulus ratio between basins, phases in the reference basin were spread relatively evenly across seasons ( $P > 0.05$ ; Rao's spacing test<sup>14</sup>;  $H_0$  = angular uniformity). Phases in the treatment basin were strongly clustered

<sup>1</sup>Integrative Biology, University of Texas, Austin, Texas 78712, USA.

( $P < 0.01$ ; Rao's spacing test). A test of equivalence in mean direction and angular dispersion between basins was negative ( $P = 0.46$  and  $P = 0.22$ ; Rao's two-sample test). Notice the strong phase shift and increase in the length of the total biomass vector in the treatment basin (Fig. 3). Although most species showed some difference in phase and amplitude between basins, the most significant impact on total biomass variation was the marked reduction in amplitude of two acid-sensitive species: *Leptodiaptomus minutus*, a herbivorous copepod, and *Diacyclops thomasi*, a predatory copepod. Both of these species reach high abundances during winter months in the reference basin, serving to counterbalance the group of summer-peaking species (see also Supplementary Fig. 1). The outcome of this counterbalance effect was to reduce seasonal biomass variation in the reference basin. Near absence of these winter dominant species in the treatment basin caused an overall reduction in total biomass and a marked increase in seasonal biomass variation. The net effect of all these changes was strongly synchronous dynamics in the disturbed system.

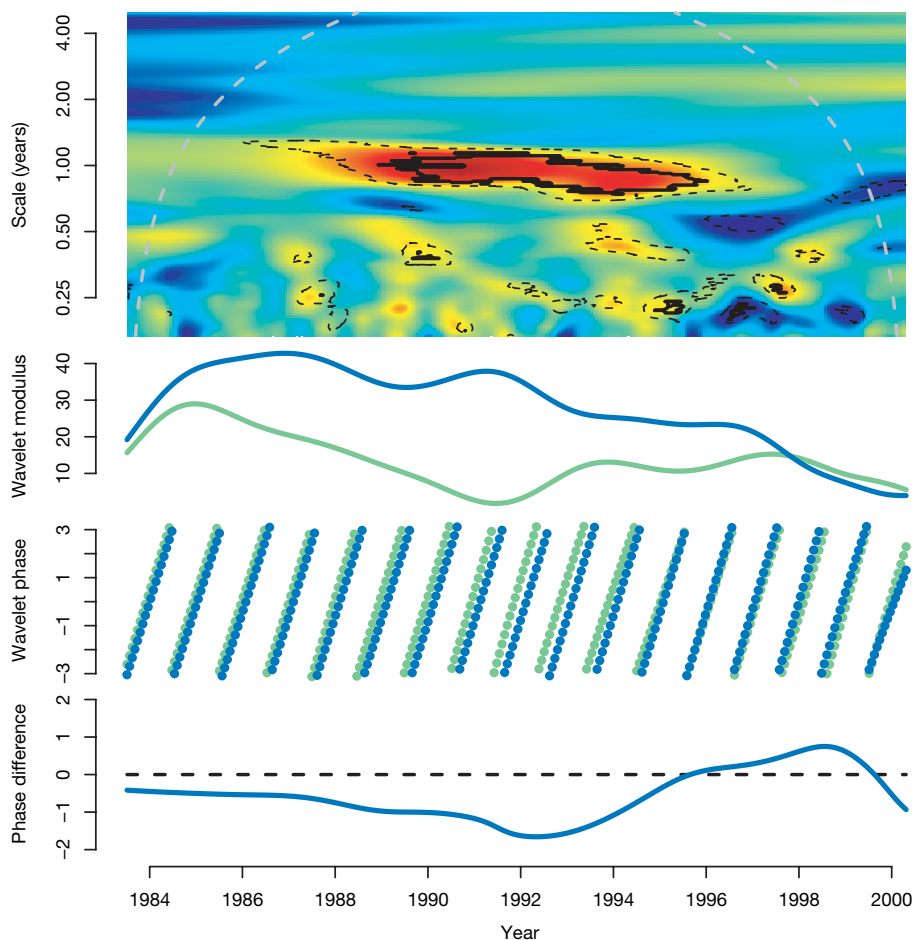
Although previous studies have found evidence of synchrony between individual species pairs in these data<sup>2,4</sup>, to my knowledge this is the first study to demonstrate conclusively that compensatory

shifts can lead to whole-community synchrony after disturbance. Notably, the observed synchrony is directly related to an increase in temporal variability of total community biomass, a clear indication of changed ecological function despite compensatory maintenance of biomass within functional groups. One may legitimately question how general these results are. There is of course a lack of treatment replication, a problem inherent in large-scale ecological experiments. A recent meta-analysis of published community data sets suggested that positive temporal covariance within communities may be more common than negative covariance<sup>15</sup>. The study only reported frequencies of the occurrence of positive and negative values and did not attempt to assert whether any of the observed covariances are larger or smaller than would be expected by chance alone. Nonetheless, the conclusion that communities can experience net positive covariance owing to external forces like seasonality is consistent with the general conclusions of this study. It is notable that the reference basin in this study showed a marginally significant trend towards compensation during some phases of the experiment, demonstrating that nonrandom negative covariance does occur under baseline conditions in at least one case.



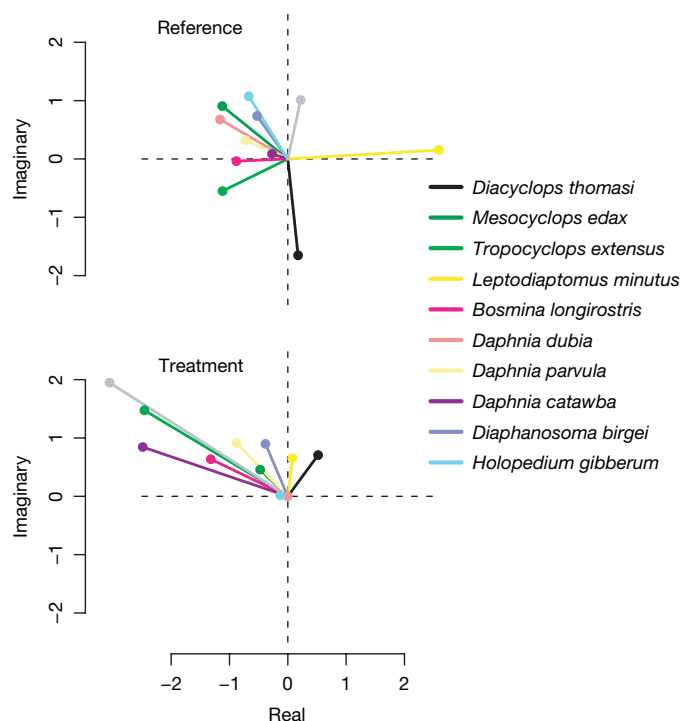
**Figure 1 | Wavelet modulus ratio for reference and treatment basins.** The top inset shows the difference in pH between the basins. The trend line is a smooth spline fit. Heat maps show the modulus ratio scaled from 0 (dark blue denotes compensation) to 1 (dark red represents synchrony). Thin dashed lines delineate regions where the modulus ratio fell outside phase-perturbed 95% confidence intervals (see Methods). Heavy solid lines indicate regions significantly different from the null model with the false

discovery rate controlled at the 5% level. The dashed grey lines indicate  $\pm 1$  unit of scale (or equivalently  $\pm 1$  cycle) of boundaries. Values nearer the boundaries should be interpreted with caution. Inset time series show biomass summed across the ten species. The right margin plots show the global modulus ratio (see Methods) with phase-perturbed 95% confidence intervals.



**Figure 2 | Modulus ratio difference between basins.** The heat map shows the difference between the basins (contours as in Fig. 1). Wavelet moduli (blue represents reduced pH; green symbolizes reference) were computed on the time series of summed biomass; larger values indicate greater summed biomass variation. Phase angles and phase difference were also computed from summed biomass.

Mechanisms explaining synchronous or compensatory dynamics in response to disturbance depend crucially on the timing, magnitude and type of disturbance. The disturbance in this study was administered as a sustained reduction in pH, but disturbances can also come as single shocks or periodic fluctuations. The results might have been quite different with periodic disturbance. The simplest explanation for community synchrony is if all species respond identically to periodic disturbance<sup>16,17</sup>. However, it seems likely that cascading ecological interactions in complex communities<sup>18–20</sup> could easily override simple linear effects. For example, a recent modelling study suggested that periodic disturbance can induce synchrony in previously asynchronous communities by disrupting lagged resource–consumer cycles<sup>21</sup>. Interestingly, the synchrony in these models was associated with a reduction in overall biomass variation, the opposite of results presented here. Another recent study showed that periodic nutrient pulsing caused an overall reduction in synchrony and a decrease in biomass variability in experimentally manipulated communities<sup>22</sup>, suggesting that periodic inputs can disrupt endogenous cycles and reduce temporal biomass variation. The candidate mechanism at work in this study was entirely different as species responded to a sustained change in environmental conditions rather than periodic disturbance. Seasonality is an important factor affecting community dynamics in the LRL ecosystem. However, seasonality alone was not sufficient to generate coherent dynamics as annual-scale synchrony was not present in the reference basin. Variation in species responses to reduced pH resulted in strong seasonal clustering and synchrony in the treatment basin. These results suggest that current theory needs to become more attuned to potential interactions between environmental change, seasonality and internal community dynamics in order to provide more accurate predictions of community-level response to large-scale disturbance. Further development of time-frequency methods for community



**Figure 3 | Snapshot of species phase relationships at the 1-year timescale.** Wavelets were centred on 10 October 1990, the approximate date of maximum difference in wavelet modulus ratio between basins (see heat map at the top of Fig. 2). Contribution of data values preceding and following the central date decay by 63% per year. Vector lengths correspond to the log-transformed modulus of each species. Vector orientations indicate species phases computed at the annual timescale. The grey vectors give the same results, but for total biomasses rather than individual species.



analysis would be welcome, as would the development of new models that can account for transient synchrony arising through complex interaction between seasonality, disturbance and population dynamics.

## METHODS SUMMARY

I used the continuous wavelet transform<sup>13,23–25</sup>

$$w_k(t, s) = s^{-1} \int_{-\infty}^{\infty} \psi\left(\frac{t-\tau}{s}\right) x_k(\tau) d\tau \quad (1)$$

to detect synchronous and compensatory community dynamics. Here,  $s$  sets the scale of analysis,  $\psi(\tau)$  is the wavelet function and  $x_k(\tau)$  is the biomass of the  $k$ th species at time  $\tau$ . I used the complex-valued Morlet wavelet<sup>26</sup>  $\psi(\tau) = \pi^{-1/4} \exp(2\pi i\tau - \frac{1}{2}\tau^2)$  in all analyses. Coherency was measured using the localized wavelet modulus ratio

$$\rho(t, s) = \frac{A_{t,s}(|\sum_k w_k(t, s)|)}{A_{t,s}(\sum_k |w_k(t, s)|)} \quad (2)$$

where  $A_{t,s}(\cdot) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{t-\tau}{s})^2} (\cdot) d\tau$  and  $|\cdot|$  denotes the modulus or complex norm. The numerator of equation (2) quantifies aggregate biomass variation at time  $t$  and scale  $s$ , whereas the denominator captures individual species variation. Under compensatory dynamics, aggregate variation is small relative to individual species variation and the modulus ratio tends towards zero. With synchrony, aggregate variation approaches the sum of individual species variation and the modulus ratio tends towards one.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 5 March; accepted 25 March 2008.**

1. Micheli, F. *et al.* The dual nature of community variability. *Oikos* **85**, 161–169 (1999).
2. Keitt, T. H. & Fischer, J. Detection of scale-specific community dynamics using wavelets. *Ecology* **87**, 2895–2904 (2006).
3. Vasseur, D. A. & Gaedke, U. Spectral analysis unmasks synchronous and compensatory dynamics in plankton communities. *Ecology* **88**, 2058–2071 (2007).
4. Fischer, J. M., Frost, T. M. & Ives, A. R. Compensatory dynamics in zooplankton community responses to acidification: Measurement and mechanisms. *Ecol. Appl.* **11**, 1060–1072 (2001).
5. Frost, T. M., Fischer, J. M., Klug, J. L., Arnott, S. E. & Montz, P. K. Trajectories of zooplankton recovery in the Little Rock Lake whole-lake acidification experiment. *Ecol. Appl.* **16**, 353–367 (2006).
6. Tilman, D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **80**, 1455–1474 (1999).
7. McCann, K. S. The diversity-stability debate. *Nature* **405**, 228–233 (2000).

8. Cottingham, K. L., Brown, B. L. & Lennon, J. T. Biodiversity may regulate the temporal variability of ecological system. *Ecol. Lett.* **4**, 72–85 (2001).
9. Loreau, M. *et al.* Ecology - Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**, 804–808 (2001).
10. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
11. Forst, T. M. *et al.* The experimental acidification of Little Rock Lake. In *Long-Term Dynamics of Lakes in the Landscape* (eds Magnuson, J. J., Kratz, T. K. & Benson, B. J.) (Oxford Univ. Press, New York, 2005).
12. Maraun, D. & Kurths, J. Cross wavelet analysis: significance testing and pitfalls. *Nonlin. Processes Geophys.* **11**, 505–514 (2004).
13. Daubechies, I. *Ten Lectures On Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 1992).
14. Rao, J. S. & SenGupta, A. *Topics in Circular Statistics* (World Scientific, Singapore, 2001).
15. Houlihan, J. E. *et al.* Compensatory dynamics are rare in natural ecological communities. *Proc. Natl Acad. Sci. USA* **104**, 3273–3277 (2007).
16. Moran, P. A. P. The statistical analysis of the Canadian lynx cycle. II. synchronization and meteorology. *Aust. J. Zool.* **1**, 291–298 (1953).
17. Kent, A. D., Yannarell, A. C., Rusak, J. A., Triplett, E. W. & McMahon, K. D. Synchrony in aquatic microbial community dynamics. *ISME J.* **1**, 38–47 (2007).
18. Carpenter, S. R., Kitchell, J. F. & Hodgson, J. R. Cascading trophic interactions and lake productivity. *Bioscience* **35**, 634–639 (1985).
19. Ives, A. R. Predicting the response of populations to environmental change. *Ecology* **76**, 926–941 (1995).
20. Pace, M. L., Cole, J. J., Carpenter, S. R. & Kitchell, J. F. Trophic cascades revealed in diverse ecosystems. *Trends Ecol. Evol.* **14**, 483–488 (1999).
21. Vasseur, D. A. & Fox, J. W. Environmental fluctuations can stabilize food web dynamics by increasing synchrony. *Ecol. Lett.* **10**, 1066–1074 (2007).
22. Downing, A. L., Brown, B. L., Perrin, E. M., Keitt, T. H. & Leibold, M. A. Environmental fluctuations induce scale-dependent compensation and increase stability in plankton ecosystems. *Ecology* (in the press).
23. Mallat, S. *A Wavelet Tour of Signal Processing* 2nd edn (Academic, New York, 1999).
24. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
25. Percival, D. B. & Walden, A. T. *Wavelet Methods in Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge Univ. Press, Cambridge, UK, 2002).
26. Grossmann, A. & Morlet, J. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* **15**, 723–736 (1984).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** I thank J. Fischer and M. Kirkpatrick for their comments on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.H.K. (tkeitt@mail.utexas.edu).

## METHODS

**Wavelet analysis.** The LRL data were collected roughly every 2 weeks in warmer months and approximately every 6 weeks in winter. Irregular sampling precludes standard numerical methods based on the fast Fourier transform. Instead, I directly computed the inner product of the data samples and the scaled and translated wavelet function. The zero-padded signal for each species is given by

$$x_k(\tau) = \sum_i \delta(\tau - v_i) x_k(v_i) \quad (3)$$

where  $v_i$  is the time of the  $i$ th sample and  $\delta(\cdot)$  is the delta function. Substitution in equation (1) yields

$$w_k(t, s) = s^{-1} \sum_i \psi\left(\frac{t - v_i}{s}\right) x_k(v_i) \quad (4)$$

where the equivalence holds under relatively weak smoothness and continuity conditions for  $x$  and  $s$  at or above the Nyquist scale<sup>13</sup>. I approximate the Nyquist scale as twice the mean time lag between samples and require  $s$  to be at least 1.5 times this value. Equation (4) was evaluated at regular intervals separated by the average time lag between successive LRL sampling dates (21 days) and  $s$  ranged from 2 months to 5 years.

**Wavelet modulus ratios and phases.** The wavelet modulus ratio can be estimated locally (equation (2)) or accumulated over the entire time series. The global case is given by

$$\rho(s) = \frac{\langle |\sum_k w_k(t, s)| \rangle}{\langle \sum_k |w_k(t, s)| \rangle} \quad (5)$$

where  $\langle \cdot \rangle$  denotes the arithmetic mean with respect to time (see margin plots in Fig. 1). Wavelet phase was computed as  $\theta = \tan^{-1}\left(\frac{a}{b}\right)$  where  $a$  and  $b$  are real and imaginary parts of the wavelet coefficient. Annual-scale phase differences were rescaled to months and further adjusted to sidestep the December to January discontinuity<sup>24</sup> using the formula  $\frac{6}{\pi} [(\Delta\theta + 3\pi) \text{ modulo } 2\pi] - 6$ .

**Null model simulation via stochastic wavelets.** A common approach to estimating a null distribution based on independent fluctuations is repeated shuffling of individual species time series<sup>4,27</sup>. Time series randomization alters serial correlations and can drastically alter the underlying null model such that estimated probabilities are biased relative to the ideal null model where species fluctuate independently, but all other properties of the original time series are preserved<sup>28</sup>. To overcome this obstacle, I developed an alternative method where random perturbations are introduced to the wavelet and the original time series remain intact. The stochastic wavelets, indexed by species and scale, are given by

$$\psi_{k,s}(\tau) = \pi^{-1/4} \exp\left(2\pi i[\tau + \varepsilon_{k,s}] - \frac{1}{2}\tau^2\right) \quad (6)$$

where  $\varepsilon_{k,s}$  are uniform samples from the interval  $(-1/2, 1/2)$ . The perturbation is equivalent to a simple phase shift in the wavelet domain for each species at each scale. Null distributions were estimated from 1,000 independent trials. I controlled the false discovery rate at the 5% level using the method previously described<sup>29</sup>, which is valid in the presence of interdependence. Application of the phase-perturbation method to simulated data indicates high statistical power and low bias in the presence of zeros in time series data (Supplementary Figs 2 and 3).

27. Ernest, S. K. M. & Brown, J. H. Homeostasis and compensation: the role of species and resources in ecosystem stability. *Ecology* **82**, 2118–2132 (2001).

28. Solow, A. & Duplisea, D. Testing for compensation in a multi-species community. *Ecosystems* **10**, 1034–1038 (2007).

29. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188 (2001).

# Control of segment number in vertebrate embryos

Céline Gomez<sup>1</sup>, Ertuğrul M. Özbudak<sup>1</sup>, Joshua Wunderlich<sup>1</sup>, Diana Baumann<sup>1</sup>, Julian Lewis<sup>2</sup> & Olivier Pourquie<sup>1,3</sup>

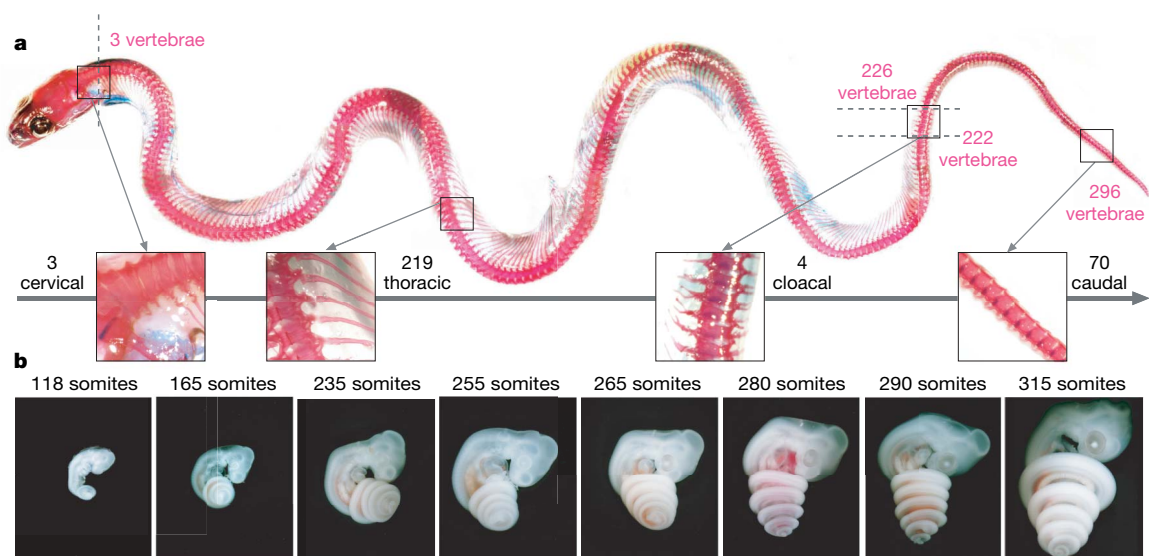
The vertebrate body axis is subdivided into repeated segments, best exemplified by the vertebrae that derive from embryonic somites. The number of somites is precisely defined for any given species but varies widely from one species to another. To determine the mechanism controlling somite number, we have compared somitogenesis in zebrafish, chicken, mouse and corn snake embryos. Here we present evidence that in all of these species a similar ‘clock-and-wavefront’<sup>1–3</sup> mechanism operates to control somitogenesis; in all of them, somitogenesis is brought to an end through a process in which the presomitic mesoderm, having first increased in size, gradually shrinks until it is exhausted, terminating somite formation. In snake embryos, however, the segmentation clock rate is much faster relative to developmental rate than in other amniotes, leading to a greatly increased number of smaller-sized somites.

Vertebrate segments are formed during early embryogenesis, when vertebrae precursors, called somites, bud off in a rhythmic fashion from the anterior part of the presomitic mesoderm (PSM). The periodic formation of somites is proposed to be controlled by a molecular oscillator—the segmentation clock—which drives the periodic activation of the Notch, Wnt and fibroblast growth factor (FGF) pathways in the PSM<sup>1,2,4</sup>. The periodic signal of the segmentation clock is converted into a repetitive series of somites by a travelling front of maturation—the wavefront or determination front—formed by a Wnt/FGF signalling gradient that regresses caudally in the PSM in concert with axis elongation<sup>5–8</sup>.

The number of somites, and hence of vertebrae, is highly variable among vertebrate species<sup>9</sup>. For instance, frogs have ~10 vertebrae, whereas humans have 33 and snakes can have more than 300. To investigate the mechanisms controlling somite numbers in vertebrates, we compared somitogenesis in the corn snake (*Pantherophis guttatus*; Fig. 1a, b), which makes a large number of somites (~315), with that in three other vertebrate species that make far fewer: zebrafish (*Danio rerio*, 31), chicken (*Gallus gallus*, 55) and mouse (*Mus musculus*, 65).

We examined the expression of the corn snake homologues of genes involved in PSM patterning and somitogenesis (Fig. 2a–l). The genes coding for fibroblast growth factor 8 (FGF8) (Fig. 2b) and its targets sprouty 2 (SPRY2; Fig. 2c) and dual specificity phosphatase 6 (DUSP6)<sup>4,10</sup> (Fig. 2d), as well as WNT3A (Fig. 2e) and its targets AXIN2 (ref. 7; Fig. 2f) and mesogenin 1 (MSGN1) (ref. 11; Fig. 2g), ephrin receptor A4 (EPHA4; Fig. 2h), the retinoic acid biosynthetic enzyme (RALDH2; Fig. 2i), paraxis (TCF15) (Fig. 2j), UNCX4.1 and MYOD (Supplementary Fig. 1) were cloned and their expression analysed by *in situ* hybridization. All of these genes (except *SPRY2*) were expressed in domains comparable to those observed in their fish or amniote counterparts<sup>4–7,12–19</sup>, supporting the existence of a Wnt/FGF posterior gradient opposing an anterior retinoic acid gradient in corn snake, as in other species.

We compared the dynamics of this gradient in snake with that in the other species. As a readout for the posterior gradients, we used *MSGN1* expression, which is controlled by the Wnt/FGF gradient

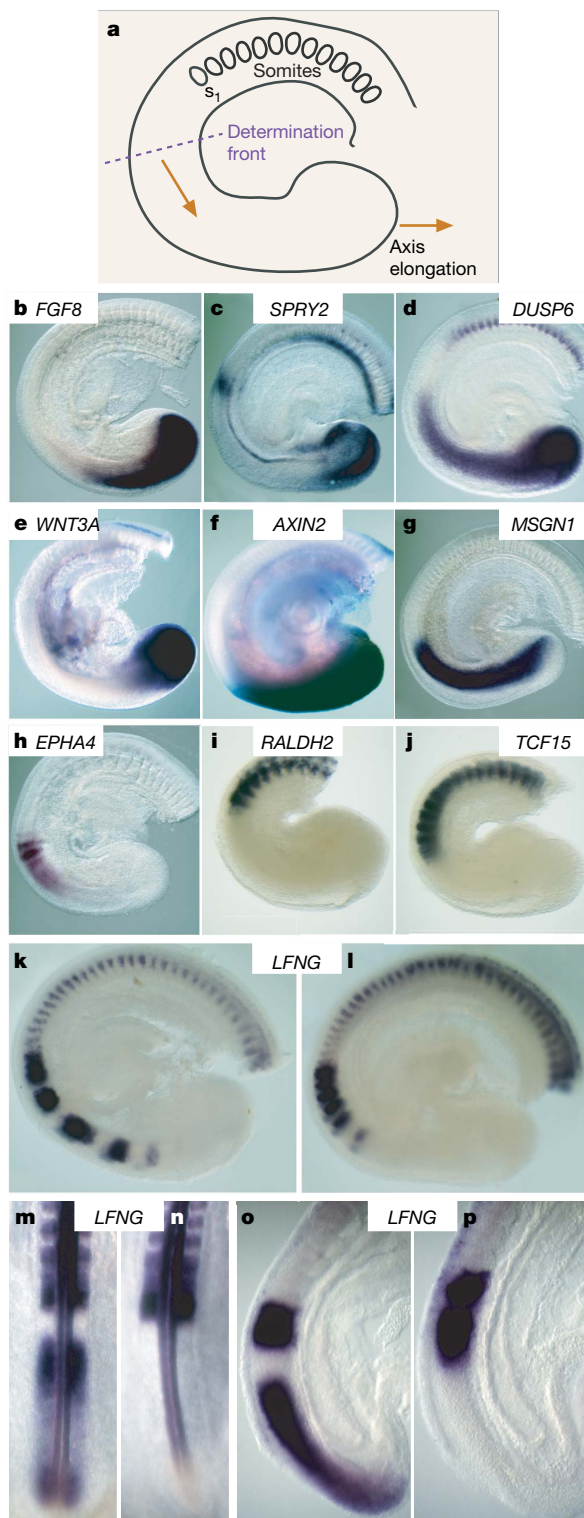


**Figure 1 | Vertebral formula and somitogenesis in the corn snake.** **a**, Alizarin staining of a corn snake showing 296 vertebrae, including 3 cervical, 219 thoracic, 4 cloacal (distinguishable by their forked

lymphapophyses) and 70 caudal. **b**, Time course of corn snake development after egg laying (118-somite embryo on the far left) until the end of somitogenesis (~315 somites).

<sup>1</sup>Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. <sup>2</sup>Vertebrate Development Laboratory, Cancer Research UK, London Research Institute, London WC2A 3PX, UK. <sup>3</sup>Howard Hughes Medical Institute, Kansas City, Missouri 64110, USA.





**Figure 2 | The corn snake determination front and segmentation clock.** **a**, Schematic drawing of a corn snake tail showing the position of the determination front (dashed line) in the PSM relative to the newly formed somite ( $s_1$ ). **b–l**, Whole-mount *in situ* hybridizations of 230-somite (**b–h, k, l**) and 260-somite (**i, j**) corn snake embryo tails: *FGF8* (**b**), *SPRY2* (**c**), *DUSP6* (**d**), *WNT3A* (**e**), *AXIN2* (**f**), *MSGN1* (**g**), *EPHA4* (**h**), *RALDH2* (**i**) and *TCF15* (**j**). Two different phases of *LFNG* expression in 230-somite corn snake embryos (**k, l**), two-day-old chicken embryos (**m, n**) and E9.5 mouse embryos (**o, p**). Lateral views in **a–l, o** and **p**. Dorsal views in **m** and **n**. Anterior to the top.

and for which the sharp anterior boundary marks the position of the mesoderm posterior 2 (*MESP2*) stripe<sup>20</sup> at the determination front<sup>3</sup> (Fig. 2a–g). We measured the regression speed of the *MSGN1* anterior boundary during somitogenesis in all four species (Fig. 3a–x) and found that it moves by one somite length during one period of somite formation, independent of the somitogenesis stage and the species (Fig. 4a and Supplementary Fig. 2). This validates an important prediction of the clock-and-wavefront model—that somite size corresponds to the distance travelled by the wavefront during one oscillation period. We plotted the ratio of *MSGN1* expression domain to PSM size as a function of stage for each species (Fig. 4b). Notably, a similar ratio was observed throughout somitogenesis in all four species, suggesting that similar processes (scaled proportionately) are occurring. Thus, the characteristics of the gradient system involved in PSM patterning seem to be conserved between corn snake and the other species examined.

We examined the cyclic gene expression associated with the amniote segmentation clock. No dynamic expression of *SPRY2*, *DUSP6* (ref. 4) or *AXIN2* (ref. 7) was evident in the snake PSM (Fig. 2c, d, f). However, *Lumatic fringe* (*LFNG*) exhibited an unexpected expression pattern that consisted of up to nine stripes of variable size and spacing in the PSM (Fig. 2k, l)<sup>21,22</sup>. Thirty-nine snake embryos were hybridized with *LFNG* and all showed a different expression pattern (Fig. 2k, l and data not shown). These data support the existence of an oscillator driving cyclic gene expression in snake embryos. The number of stripes of *LFNG* expression in corn snake is, however, several times larger than in other vertebrate species (Fig. 2k–p), suggesting that the segmentation clock might be regulated in a different way.

Assuming that one somite is formed during each clock oscillation cycle, we can deduce the period of the clock oscillations by counting somite numbers in embryos from the same clutch at various incubation times (Supplementary Methods). In the corn snake, the average somite formation rate is one pair every 100 min, compared to rates of one pair every 30, 90 and 120 min in zebrafish, chicken and mouse, respectively. The somite formation rate was found to be one pair every ~60 min in the house snake (*Lamprophis fuliginosus*) embryos, which has very similar developmental characteristics and *LFNG* expression pattern to the corn snake (Supplementary Methods and Supplementary Fig. 3).

To appreciate the significance of these periods, we need to compare them to the general rate of development, which differs between species. Comparison of the time required to reach conserved morphological landmarks (Supplementary Fig. 4) suggests that the development rate is at least three times slower in corn snake than in chicken. We also examined the lizard *Aspidoscelis uniparens*, which has the same slow general development rate as the snake<sup>23</sup> (Supplementary Methods). This species makes only ~90 somites and has a much longer somite formation time (~4 h). Therefore, relative to the development rate, the clock ticks much faster in snake than in chicken or lizard embryos.

Analysing the relationship between somitogenesis rate and growth rate of the PSM tissue confirms this hypothesis. According to the clock-and-wavefront model, each somite consists of the cells emerging from the PSM in one clock cycle. This must equal the quantity of new cells generated in the PSM by growth, at least when the PSM maintains a steady size. Thus, the somite size as a fraction of the PSM size directly reflects the duration of the clock cycle as a fraction of the average PSM cell generation time (that is, average cell-cycle time; Supplementary Box 1). In snake, this fraction is approximately one-quarter of the value observed in other species (Fig. 4c). Thus, the snake segmentation clock runs approximately four times faster relative to the average PSM cell generation time than in the other species.

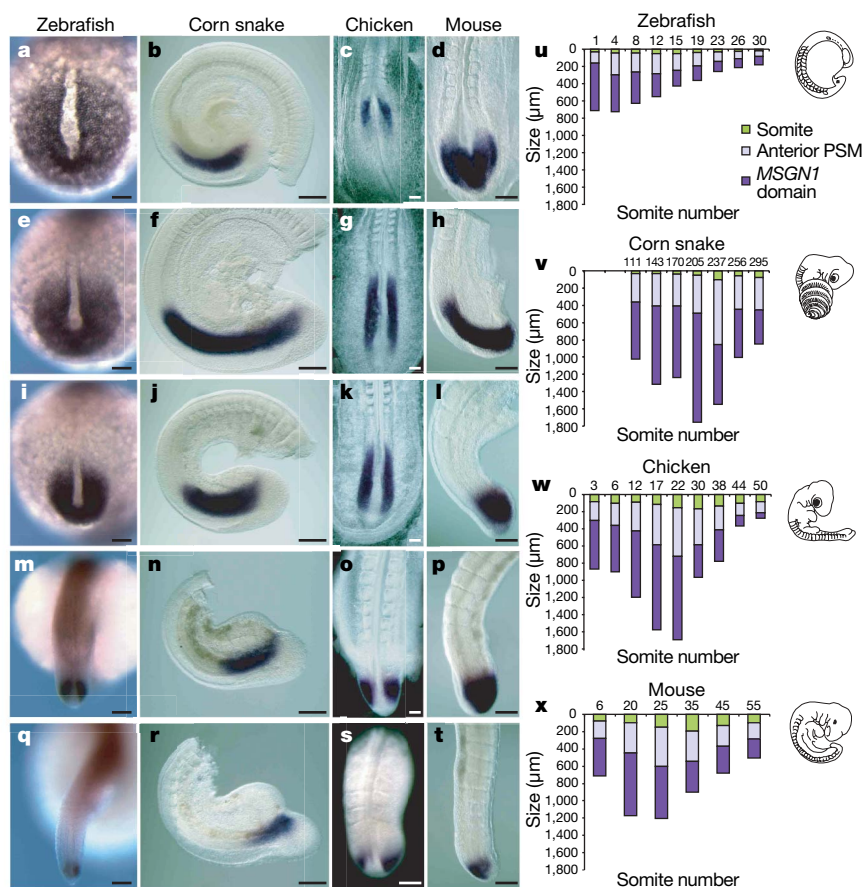
We estimated the average cell generation time and the total number of PSM cell generations required to produce a complete set of somites. For this, we took into account the way in which the PSM size

changes over the course of somitogenesis (Figs 3a–x and 4d). In zebrafish, the PSM length decreases from the beginning of somitogenesis; in amniote embryos, the PSM first increases and then decreases in size until somitogenesis ends<sup>24</sup> (Fig. 4d). From the detailed measurements of PSM length and the size of the most recently formed somite  $s_1$  as a function of developmental stage (Fig. 4d, e), combined with a knowledge of the period of the segmentation clock and the total number of somites formed, we can estimate the average cell-generation time and the total number of PSM cell generations required to generate the full set of somites (Supplementary Box 1). This calculation provides an average cell generation time in the PSM that is much longer in corn snake (~24.4 h) than in mouse (~8 h), chicken (~6.3 h) or zebrafish (~5.5 h). Direct cell-cycle measurements, using 5-bromodeoxyuridine (BrdU) incorporation followed by flow cytometry, gave a cycle time of ~34 h in corn snake embryonic tail and ~30 h in lizard, compared to ~9 h in chicken (as measured using tritiated thymidine<sup>25</sup>; Supplementary Table 4, Fig. 4f and Supplementary Fig. 5), confirming that the cell generation time is almost fourfold slower in snake and lizard than in chicken. Remarkably, the calculated number of cell generations required to generate the 315 somites in the snake (~21 generations) is only slightly greater than for the 65 somites in the mouse (~17 generations) or 55 somites in the chicken (~13 generations), although much larger than for the 31 somites in the zebrafish (~2.8 generations). Therefore, the exceptionally large number of somites in the snake, compared with that in other amniotes, is not primarily the result of a large number of generations

of PSM growth, but reflects a clock rate that is rapid in relation to the cell-cycle rate in the elongating axis.

Finally, we investigated the basis for the unusually large number of *LFNG* stripes observed in snake embryos. The number of stripes of expression reflects the number of clock cycles by which the cells at the anterior end of the PSM lag behind those in the posterior PSM. By measuring the stripe spacing, one can deduce how the clock rate changes with position in the PSM<sup>26</sup>. Using measurements of *LFNG* stripe spacing in corn snake embryos, we obtained a graph depicting the slowing down of gene oscillations in the PSM. The graphs for snake and zebrafish are almost identical (Fig. 4g), suggesting that the mechanism controlling the slowing down of cyclic gene oscillations is similar in the two species. The same manner of slowing down of gene oscillations in snake and zebrafish entails very different numbers of PSM stripes simply because of the different ratios of oscillator rate to growth rate (Supplementary Box 2). The faster the oscillator runs relative to PSM growth rate, the more stripes of cyclic gene expression are observed in the PSM.

Thus, our data show that the basic clock-and-wavefront mechanism operates according to similar principles in snake, chicken, mouse and zebrafish. In all four species, somitogenesis ends with a progressive shrinking of the PSM. This shrinking presumably reflects a gradual extinction of the signals that maintain the PSM character of cells at the tail end of the embryo<sup>27</sup>. In chicken and mouse embryos, termination of somitogenesis is an active process associated with extensive cell death in the tail bud<sup>28</sup>. Apoptosis of tail bud cells leading to axis truncation can be induced by retinoic acid treatment<sup>29</sup>;



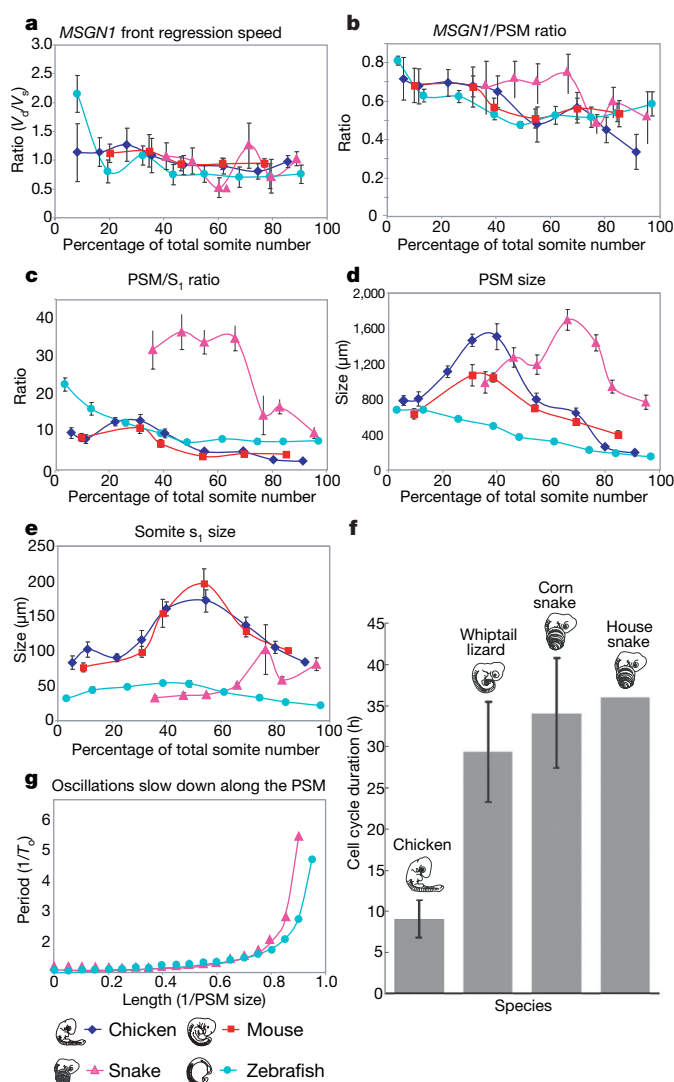
**Figure 3 | Dynamics of the PSM size in zebrafish, corn snake, chicken and mouse.** a–t, Developmental series of four vertebrate species hybridized with *MSGN1* in whole mount. Zebrafish embryos are shown at 1 (a), 8 (e), 12 (i), 19 (m) and 30 (q) somites; corn snake embryos at 165 (b), 202 (f), 251 (j), 291 (n) and 310 (r) somites; Chicken embryos at 6 (c), 17 (g), 22 (k), 30 (o) and 44 (s) somites; and mouse embryos at 6 (d), 20 (h), 35 (l), 45 (p) and

55 (t) somites. Dorsal views (a, e, i, m, q, c, g, k, o, s, d) and lateral views (b, f, j, n, r, h, l, p, t). Anterior to the top. Scale bars correspond to 100 μm (zebrafish) and 200 μm (corn snake, chicken and mouse). u–x, Graphs showing the time evolution (in somite numbers) of the PSM (blue), of the somite  $s_1$  (green) and the *MSGN1* domain (dark blue) sizes in chicken, corn snake, mouse and zebrafish embryos.



thus, it is possible that the proximity of the anterior retinoic acid domain to the tail bud caused by PSM shrinking triggers the death of paraxial mesoderm precursors. Therefore, PSM shrinking could explain the arrest of axis elongation and termination of somitogenesis. In zebrafish, the switch to PSM shrinkage occurs very early in comparison with that in amniotes, and this correlates with a small somite number. In amniotes, the switch to PSM shrinkage occurs at the transition between the trunk and the tail region, suggesting that it might be under the control of regional regulators such as *HOX* genes.

Our modelling suggests that the number of PSM generations for which somitogenesis continues changes relatively little among amniotes. In contrast, over the course of evolution there have been large changes in the ratio of the segmentation clock rate to the developmental growth rate. Variation in this ratio accounts for much of the evolutionary divergence in amniote segment number.



**Figure 4 | Comparison of somitogenesis parameters.** **a**, Ratio of the speed of the *MSGN1* front regression (in  $\mu\text{m}$  per period,  $V_d$ ) to the speed of somitogenesis (in  $\mu\text{m}$  per period,  $V_s$ ). **b**, Variation of the ratio of the *MSGN1* domain size to the PSM size during somitogenesis. **c**, Variation of the ratio of PSM size to  $s_1$  size during somitogenesis. **d**, Dynamics of the PSM size during somitogenesis. **e**, Dynamics of  $s_1$  size during somitogenesis. **f**, Graph comparing the cell-cycle time in four amniote species. Error bars represent standard deviation. Sample sizes are provided in Supplementary Methods (a–e) and in Supplementary Table 4 (f). **g**, Slowing down of the oscillation period along the PSM ( $T_0$ , clock period at the tip of the tail).

## METHODS SUMMARY

In total, 192 corn snake eggs, 13 house snake eggs and 34 whiptail lizard eggs were used in this study. Corn snake genes were cloned by PCR using standard protocols (Supplementary Table 1). Orthology was established by calculating the percentage of similarity with orthologous genes in other vertebrate species using Vector NTI (Supplementary Table 2). Whole-mount *in situ* hybridization was performed as described<sup>30</sup> with a hybridization temperature of 58 °C (Supplementary Table 3). For measurements, embryos were hybridized with a *MSGN1* probe in whole mount and were photographed. The size in micrometres of the PSM, the somite  $s_1$  and the *MSGN1* domain were measured using Zeiss LSM image browser software. Embryos were pooled in groups of five on the basis of their somite number. Measurements corresponding to each pool were averaged, and the standard deviation calculated. For the calculation of the slowing down of the period along the PSM in corn snake, interstripe distance was measured using the Zeiss LSM software in two-day post-oviposition embryos ( $n=23$ ) stained with *LFNG*, and calculations were performed as described<sup>26</sup>. Cell cycle length was measured using BrdU incorporation followed by flow cytometry analysis. Alizarin staining was performed according to standard procedures.

Received 4 October 2007; accepted 21 April 2008.

Published online 18 June 2008.

- Cooke, J. & Zeeman, E. C. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theor. Biol.* **58**, 455–476 (1976).
- Palmeirim, I., Henrique, D., Ish-Horowicz, D. & Pourquie, O. Avian *hairy* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
- Dequeant, M. L. & Pourquie, O. Segmental patterning of the vertebrate embryonic axis. *Nature Rev. Genet.* **9**, 370–382 (2008).
- Dequeant, M. L. *et al.* A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* **314**, 1595–1598 (2006).
- Dubrulle, J., McGrew, M. J. & Pourquie, O. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal *Hox* gene activation. *Cell* **106**, 219–232 (2001).
- Sawada, A. *et al.* *Fgf*/MAPK signalling is a crucial positional cue in somite boundary formation. *Development* **128**, 4873–4880 (2001).
- Aulehla, A. *et al.* *Wnt3a* plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell* **4**, 395–406 (2003).
- Aulehla, A. *et al.* A  $\beta$ -catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nature Cell Biol.* **10**, 186–193 (2008).
- Richardson, M. K., Allen, S. P., Wright, G. M., Raynaud, A. & Hanken, J. Somite number and vertebrate evolution. *Development* **125**, 151–160 (1998).
- Delfini, M. C., Dubrulle, J., Malapert, P., Chal, J. & Pourquie, O. Control of the segmentation process by graded MAPK/ERK activation in the chick embryo. *Proc. Natl Acad. Sci. USA* **102**, 11343–11348 (2005).
- Wittler, L. *et al.* Expression of *Msn1* in the presomitic mesoderm is controlled by synergism of WNT signalling and *Tbx6*. *EMBO Rep.* **8**, 784–789 (2007).
- Nakajima, Y., Morimoto, M., Takahashi, Y., Koseki, H. & Saga, Y. Identification of *Epha4* enhancer required for segmental expression and the regulation by *Mesp2*. *Development* **133**, 2517–2525 (2006).
- Niederrreither, K., Subbarayan, V., Dolle, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature Genet.* **21**, 444–448 (1999).
- Diez del Corral, R. *et al.* Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79 (2003).
- Burgess, R., Cserjesi, P., Ligon, K. L. & Olson, E. N. Paraxis: a basic helix–loop–helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol.* **168**, 296–306 (1995).
- Mansouri, A. *et al.* Paired-related murine homeobox gene expressed in the developing sclerotome, kidney, and nervous system. *Dev. Dyn.* **210**, 53–65 (1997).
- Yoon, J. K. & Wold, B. The bHLH regulator pMesogenin1 is required for maturation and segmentation of paraxial mesoderm. *Genes Dev.* **14**, 3204–3214 (2000).
- Sassoon, D. *et al.* Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature* **341**, 303–307 (1989).
- Pownall, M. E. & Emerson, C. P. J. Sequential activation of three myogenic regulatory genes during somite morphogenesis in quail embryos. *Dev. Biol.* **151**, 67–79 (1992).
- Yoon, J. K., Moon, R. T. & Wold, B. The bHLH class protein pMesogenin1 can specify paraxial mesoderm phenotypes. *Dev. Biol.* **222**, 376–391 (2000).
- McGrew, M. J., Dale, J. K., Fraboulet, S. & Pourquie, O. The *lunatic fringe* gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr. Biol.* **8**, 979–982 (1998).
- Forsberg, H., Crozet, F. & Brown, N. A. Waves of mouse *Lunatic fringe* expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. *Curr. Biol.* **8**, 1027–1030 (1998).
- Zug, G. R., Vitt, L. J. & Caldwell, J. P. *Herpetology: an Introductory Biology of Amphibians and Reptiles* 2nd edn (Academic, San Diego, 2001).
- Tam, P. P. The control of somitogenesis in mouse embryos. *J. Embryol. Exp. Morphol.* **65** (Suppl.), 103–128 (1981).



25. Primmatt, D. R., Norris, W. E., Carlson, G. J., Keynes, R. J. & Stern, C. D. Periodic segmental anomalies induced by heat shock in the chicken embryo are associated with the cell cycle. *Development* **105**, 119–130 (1989).
26. Giudicelli, F., Ozbudak, E. M., Wright, G. J. & Lewis, J. Setting the tempo in development: an investigation of the zebrafish somite clock mechanism. *PLoS Biol.* **5**, e150 (2007).
27. Cambray, N. & Wilson, V. Two distinct sources for a population of maturing axial progenitors. *Development* **134**, 2829–2840 (2007).
28. Sanders, E. J., Khare, M. K., Ooi, V. C. & Bellairs, R. An experimental and morphological analysis of the tail bud mesenchyme of the chicken embryo. *Anat. Embryol. (Berl.)* **174**, 179–185 (1986).
29. Shum, A. S. *et al.* Retinoic acid induces down-regulation of *Wnt-3a*, apoptosis and diversion of tail bud cells to a neural fate in the mouse embryo. *Mech. Dev.* **84**, 17–30 (1999).
30. Henrique, D. *et al.* Expression of a *Delta* homologue in prospective neurons in the chicken. *Nature* **375**, 787–790 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The authors thank M. Gibson, B. Rubinstein, P. Francois and members of the Pourqu  laboratory for critical reading and discussions, M. Wahl

for the mouse *LFNG* pictures, members of the Reptile and Aquatics Department, J. Chatfield for editorial assistance, and S. Esteban for artwork. Research was supported by Stowers Institute for Medical Research, and in part by a Defense Advanced Research Projects Agency (DARPA) grant (O.P.). J.L. is supported by Cancer Research UK. Zebrafish were obtained from the Zebrafish International Resource Center (ZIRC) at the University of Oregon, which is supported by a grant from the NIH-NCRR. O.P. is a Howard Hughes Medical Institute Investigator.

**Author Contributions** C.G. and O.P. designed the experiments, C.G. cloned the snake genes and performed the mouse, chicken and snake *in situ* hybridizations, E.M.O. performed the fish *in situ*, C.G. and E.M.O. performed the measurements and analysed the data with O.P. D.B. established the corn snake and zebrafish colony and produced the embryos. C.G. and J.W. performed the cell cycle analysis. J.L. performed the mathematical modelling. C.G., E.M.O., J.L. and O.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** Sequences of genes described in this paper have been deposited into GenBank under accession numbers EU196456, EU196465, EU232010, EU196457, EU196458, EU196459, EU196460, EU196466, EU196461, EU196464, EU196462 and EU196463. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to O.P. ([olp@stowers-institute.org](mailto:olp@stowers-institute.org)).

## LETTERS

# The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards

Kathryn A. Burke<sup>1</sup>, Theresa M. Franz<sup>2</sup>, Danielle N. Miller<sup>4</sup> & Geoffrey Schoenbaum<sup>2,3</sup>

Cues that reliably predict rewards trigger the thoughts and emotions normally evoked by those rewards. Humans and other animals will work, often quite hard, for these cues. This is termed conditioned reinforcement. The ability to use conditioned reinforcers to guide our behaviour is normally beneficial; however, it can go awry. For example, corporate icons, such as McDonald's Golden Arches, influence consumer behaviour in powerful and sometimes surprising ways<sup>1</sup>, and drug-associated cues trigger relapse to drug seeking in addicts and animals exposed to addictive drugs, even after abstinence or extinction<sup>2,3</sup>. Yet, despite their prevalence, it is not known how conditioned reinforcers control human or other animal behaviour. One possibility is that they act through the use of the specific rewards they predict; alternatively, they could control behaviour directly by activating emotions that are independent of any specific reward. In other words, the Golden Arches may drive business because they evoke thoughts of hamburgers and fries, or instead, may be effective because they also evoke feelings of hunger or happiness. Moreover, different brain circuits could support conditioned reinforcement mediated by thoughts of specific outcomes versus more general affective information. Here we have attempted to address these questions in rats. Rats were trained to learn that different cues predicted different rewards using specialized conditioning procedures that controlled whether the cues evoked thoughts of specific outcomes or general affective representations common to different outcomes. Subsequently, these rats were given the opportunity to press levers to obtain short and otherwise unrewarded presentations of these cues. We found that rats were willing to work for cues that evoked either outcome-specific or general affective representations. Furthermore the orbitofrontal cortex, a prefrontal region important for adaptive decision-making<sup>4</sup>, was critical for the former but not for the latter form of conditioned reinforcement.

Rats ( $n = 13$ ) received neurotoxic lesions of the orbitofrontal cortex (OFC; see Supplementary Information for lesion drawings); controls ( $n = 19$ ) received either sham surgeries with saline infusions into the OFC or no surgery. Sham and non-surgical controls showed no differences in subsequent measures of conditioning, thus they were combined into a single group. After surgery, all rats were trained in a specialized pavlovian conditioning task, termed transreinforcer blocking<sup>5,6</sup>. Blocking refers to the observation that a cue that predicts reward will prevent the formation of associations between that reward and any other cues that are present (see Supplementary Discussion for details on alternative interpretations of blocking). Thus, if a rat is trained that the illumination of a light predicts food and is later presented with that same light and a tone followed by the same food, the rat will not learn to associate the tone with any of the information evoked by the presentation of that food (for example, its

sensory qualities and general affective properties). The light prevents, or blocks, rats from forming associations between the tone and the food.

The form of transreinforcer blocking that we have used varies this procedure by using two different but equally preferred outcomes. The light is initially presented alone followed by one outcome. Subsequently, the light is presented in compound with the tone, followed by the second outcome. Because both outcomes are equally preferred, they trigger comparable emotional responses<sup>5</sup>. These general affective properties are already predicted by the light, thus the tone is blocked from forming associations with them. However, the light does not predict information that is specific to the second outcome, such as its particular reinforcing sensory properties. Thus, the tone is able to form associations with these outcome-specific properties. As a result, the tone evokes thoughts of a specific outcome without triggering general affective representations. Evidence to support this comes from studies showing that cues trained this way facilitate responses leading to the same outcome, thought to be mediated through the activation of a neural representation of that outcome and its specific features. These cues, however, fail to facilitate responses that lead to different but similarly valenced outcomes, thought to be mediated by neural representations of their shared affective properties<sup>5</sup> (see Supplementary Discussion for additional details). Therefore, transreinforcer blocking creates a cue that can be used to test whether conditioned reinforcement is mediated by outcome-specific information.

Our implementation of transreinforcer blocking is illustrated in Fig. 1 (see Methods for further details). Rats received 12 days of conditioning in which two unique visual cues, A and B (cue light or house light, counterbalanced), were paired with two unique outcomes, O1 and O2 (grape- or banana-flavoured sucrose pellets, counterbalanced), respectively. These pellets were equally preferred but could be discriminated and their value independently manipulated; there was no effect of lesion on these preferences (see Supplementary Results). Acquisition of pavlovian conditioned responding is shown in Fig. 1a. Both control and lesioned rats learned to respond to both cues, and there were no differences in their rates of responding. A three-way analysis of variance (ANOVA) test (lesion  $\times$  cue  $\times$  day) showed a significant main effect of day ( $F_{11,330} = 9.54$ ,  $P < 0.001$ ) but no other main effects nor any interactions with either cue or lesion ( $F < 1.41$ ,  $P > 0.16$  in all cases).

After the initial conditioning, the rats underwent 9 days of compound conditioning in which they received presentations of the two 'fully conditioned' cues, A and B, in compound with two new cues, termed X and Y (tone and white noise, counterbalanced). AX led to delivery of O1, which was the same outcome predicted by A alone. Because A already predicted this outcome, X was 'blocked' from becoming associated with any of the information evoked by O1.

<sup>1</sup>Program in Neuroscience, <sup>2</sup>Department of Anatomy and Neurobiology, <sup>3</sup>Department of Psychiatry, University of Maryland School of Medicine, 20 Penn Street, HSF-2 S251 Baltimore, Maryland 21201, USA. <sup>4</sup>The Ingenuity Project, Baltimore Polytechnic Institute, 1400 West Coldspring Lane, Baltimore, Maryland 21209, USA.

BY also led to presentation of O1; however, O1 was a different outcome than that predicted by B alone. As a result, Y was 'partially conditioned', becoming associated with properties unique to O1, including its specific sensory properties, while remaining blocked from becoming associated with the general affective properties shared by O1 and O2. Conditioned responding during compound training is shown in Fig. 1b. Both control and lesioned rats continued to respond to the cues in this phase, and there were no differences in their rates of responding. A three-way ANOVA (lesion  $\times$  cue  $\times$  day) showed neither a main effect of lesion nor any interactions with lesion ( $F < 1.32$ ,  $P > 0.23$  in all cases).

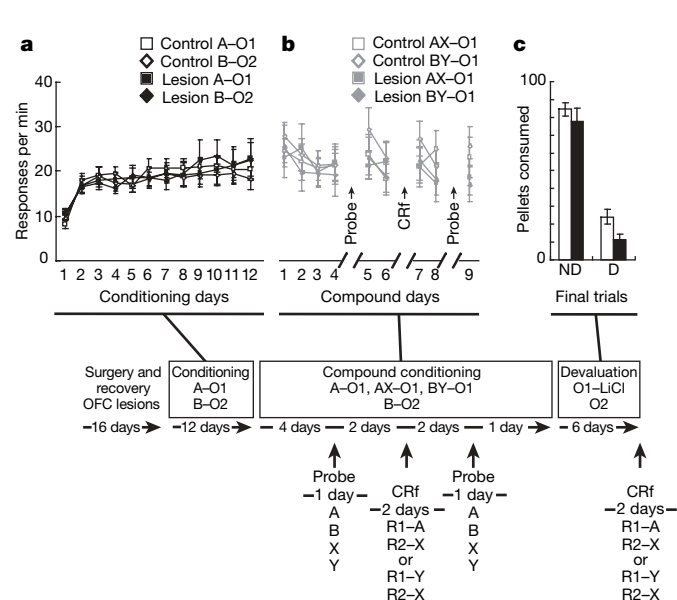
As part of this compound conditioning, we tested the ability of the pavlovian cues to serve as conditioned reinforcers. For this, control and lesioned rats were each divided into two groups such that there were no differences in conditioned responding between them. Rats in each group received two 30-min training sessions in which two levers were made available in the training chambers. For group A, pressing one lever resulted in a 1-s presentation of A, and pressing the other resulted in a 1-s presentation of X. For group Y, pressing one lever resulted in a 1-s presentation of Y, and the other resulted in a 1-s presentation of X.

The average number of responses (presses) on each lever for rats in groups A and Y is shown in Fig. 2a and b, respectively. Control rats responded similarly for A and Y; lesioned rats responded for A, but not for Y (see Supplementary Results for additional responding for B in sham and OFC-lesioned rats). A three-way ANOVA (lesion  $\times$  A/Y group  $\times$  lever) showed main effects of lever ( $F_{1,28} = 47.3$ ,  $P < 0.0001$ ), group ( $F_{1,28} = 20.1$ ,  $P < 0.001$ ) and lesion ( $F_{1,28} = 4.437$ ,  $P < 0.044$ ), and significant interactions between lesion and group ( $F_{1,28} = 16.427$ ,  $P < 0.001$ ) and between lesion,

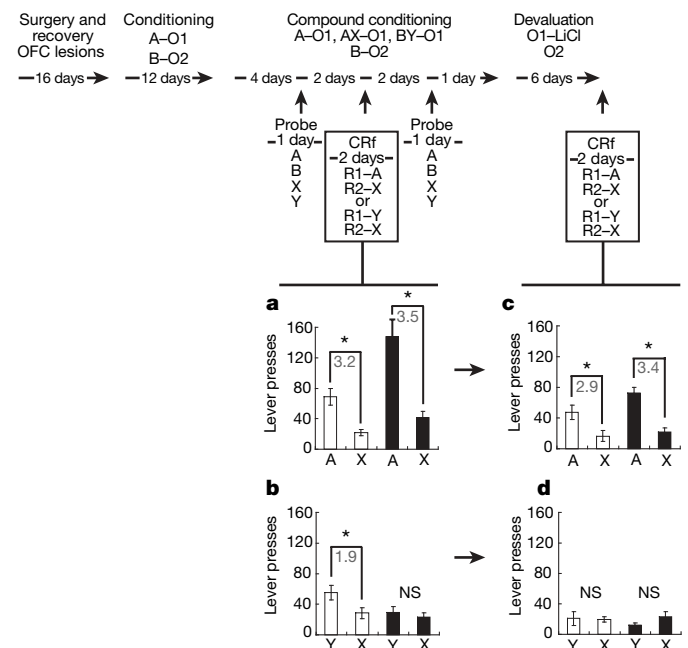
group and lever ( $F_{1,28} = 8.582$ ,  $P < 0.007$ ). Planned post-hoc tests showed significant differences in responding for A versus X in control and lesioned rats and for Y versus X in controls ( $P < 0.05$ ), but there was no significant difference in responding for Y versus X in lesioned rats ( $P = 0.7$ ).

These results demonstrate that conditioned reinforcement in rats can be mediated by outcome-specific representations. Indeed, a comparison of response ratios (A/X versus Y/X in controls, given in Fig. 2) demonstrated no significant difference in the strength of the conditioned responding between A and Y in controls ( $P = 0.18$ ). Thus, outcome-specific information was sufficient for normal responding in our task. Furthermore, because OFC-lesioned rats were impaired only when responding for Y (compare the lesion groups in Fig. 2a and b), these results show that the OFC is critical for conditioned reinforcement only when responding is mediated by outcome-specific information.

Although lesioned rats failed to respond to Y, when tested for A they responded more than controls. This increased responding was not specific to A. Unplanned post-hoc tests showed significant differences in responding in lesioned versus control rats for A ( $P = 0.0002$ ) and for X ( $P = 0.034$ ). Moreover, the increased responding for each was proportional, thus a comparison of response ratios (A/X in controls and lesioned rats, given in Fig. 2) resulted in no significant difference between controls and lesioned rats ( $P = 0.40$ ). Although a decline in responding to A in lesioned rats owing to the loss of a contribution from outcome-specific information might have been expected, general increases have been seen in other studies on the effects of prefrontal, including orbitofrontal, damage on conditioned reinforcement<sup>7,8</sup>. It is possible that affective



**Figure 1 | Effect of orbitofrontal lesions on pavlovian conditioning and reinforcer devaluation.** Shown is the experimental timeline, with boxes linking conditioning, compound conditioning, and reinforcer devaluation phases to data from each phase. In the timeline and figures, A, B, X and Y are training cues; R1 and R2 are instrumental responses; and O1 and O2 are different flavoured sucrose pellet reinforcers. **a**, Responses per minute in the food cup during the first period of the conditioned stimulus (CS) (first 8 s) for A and B during pavlovian conditioning sessions. There were no effects of lesions. **b**, Responses per minute in the food cup during the first cue period, for AX and BY during compound pavlovian conditioning sessions. The timing of the two probe tests and the pre-devaluation conditioned reinforcement (CRf) testing are indicated by arrows. There were no effects of lesions. **c**, Food pellets consumed during the final two days of devaluation (days 5 and 6) for control (open bar) and lesion (filled bar) groups. There were no effects of lesions. D, devalued pellet; ND, non-devalued pellet. Error bars denote s.d.



**Figure 2 | Effect of orbitofrontal lesions on conditioned reinforcement for a fully conditioned A cue, for a blocked X cue, and for the partially blocked Y cue before and after reinforcer devaluation.** Shown is the experimental timeline linked to data from each conditioned reinforcement (CRf) test. In the timeline and figures, A, B, X and Y are training cues; R1 and R2 are instrumental responses; and O1 and O2 are different flavoured sucrose pellet reinforcers. **a–d**, Lever pressing for A versus X, or Y versus X, in control (open bars) and lesioned (filled bars) rats before (**a**, **b**) and after (**c**, **d**) devaluation. Lesions diminished responding for Y before devaluation (**a**, **b**), controls diminished lever pressing for Y after devaluation (**c**, **d**). Lever pressing is averaged across two 30-min sessions in each figure. Asterisks indicate significance at  $P < 0.05$  on post-hoc contrast testing; the grey numbers indicate the ratio of responding on the two levers for each significant comparison. NS, not significant. Error bars denote s.d.



information, which is still mobilized by A in lesioned rats, is able to compensate for or take over from the loss of the outcome information and even generalizes to activate behaviour on the other lever. This idea is supported by the fact that reinforcer devaluation, which should also preferentially affect behaviour mediated by outcome representations, fails to affect conditioned reinforcement<sup>9</sup>. Indeed an inability to signal the expected outcome may actually contribute to this over-responding, because it might impair extinction of the conditioned reinforcer's affective value when primary reward is not presented after the cue.

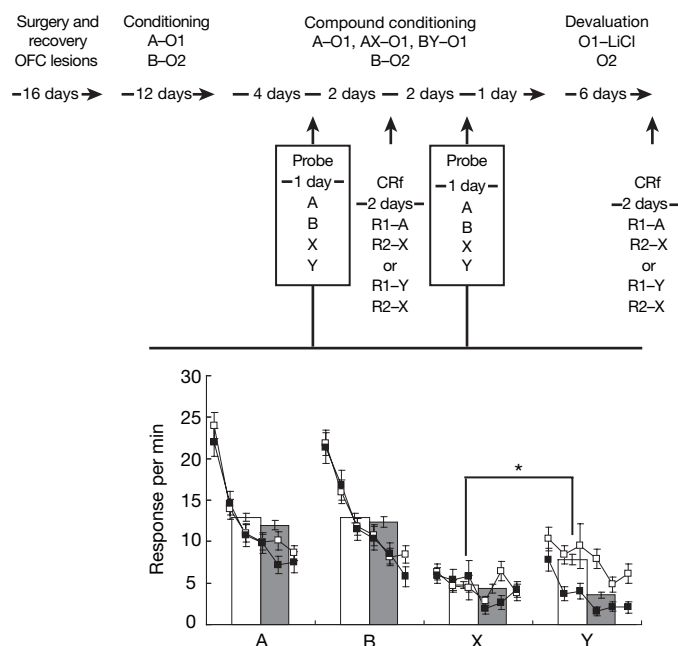
To confirm that responding for Y (but not A) was dependent on the value of the specific outcome predicted by these cues, we reassessed the ability of the cues to support conditioned reinforcement after devaluation of O1. Devaluation was conducted over 6 days. On days 1, 3 and 5, the rats received 100 pellets of the O1 outcome followed by illness, induced by lithium chloride injection. On days 2, 4 and 6, rats received the same amount of O2 without illness. As illustrated in Fig. 1c, devaluation caused a significant and selective reduction in consumption of O1 in both groups. ANOVA (lesion  $\times$  outcome) showed that both control and lesioned rats consumed significantly less of O1 than O2 on the final day of testing ( $F_{1,30} = 311.07$ ,  $P < 0.0001$ ), and there were no effects of lesion on this devaluation effect ( $F < 3.59$ ,  $P > 0.068$ , in all cases). There were also no effects of flavour (see Supplementary Results).

After devaluation, rats were returned to the training chambers with the levers available. Locations and programmed consequences remained unchanged. Average lever pressing over two 30-min sessions is shown in Fig. 2c, d. Devaluation diminished conditioned reinforcement for Y in controls but had little apparent effect on conditioned reinforcement for A in either controls or lesioned rats. A three-way ANOVA (lesion  $\times$  A/Y group  $\times$  lever) resulted in main effects of lever ( $F_{1,28} = 10.02$ ,  $P < 0.004$ ) and group ( $F_{1,28} = 6.675$ ,  $P < 0.015$ ), as well as interaction between group and lever ( $F_{1,28} = 14.237$ ,  $P < 0.001$ ). However, there were no longer any main effects or any interactions with lesion ( $F < 1.73$ ,  $P > 0.19$  in all cases). Planned post-hoc tests showed significant differences in responding

in both control and lesioned rats for A versus X ( $P < 0.05$  in both cases) but not for Y versus X ( $P > 0.50$  in both cases). Comparison of response ratios across devaluation demonstrated no significant change in the strength of conditioned responding for A in either controls ( $P = 0.37$ ) or lesioned rats ( $P = 0.7$ ).

These results demonstrate that conditioned responding in controls for Y is mediated by information about the value of the specific outcome that Y predicts, because control rats stopped responding for Y when that outcome was devalued (Fig. 2d, control group). The fact that the OFC would be critical for responding for Y brings the role of the OFC in conditioned reinforcement into accord with a growing consensus that the OFC is critical for cue-evoked behaviours when those behaviours depend critically on the value of the specific outcome that the cue predicts<sup>4</sup>. Second, these results confirm previous reports that conditioned reinforcement can be mediated by the ability of cues to trigger devaluation-insensitive general affective representations, because rats continued to respond for A even after devaluation (control group in Fig. 2c)<sup>9</sup>. The OFC does not seem to be necessary for this form of conditioned reinforcement (lesion group in Fig. 2c).

Notably, the effects of OFC lesions on conditioned reinforcement paralleled the effects of these lesions on pavlovian responding to these cues, assessed by presenting the rats with each cue (A, B, X and Y) individually in two extinction tests. These probe tests were conducted before and after the first conditioned reinforcement test. Each cue was presented six times without reinforcement. Average conditioned responding across trials in these extinction sessions is shown in Fig. 3. Both control and lesioned rats showed conditioned responding to cues A and B, whereas control but not lesioned rats acquired conditioned responding to Y. Neither group acquired a conditioned response to X. Consistent with this description, a three-way ANOVA (lesion  $\times$  cue  $\times$  trial) showed main effects of cue, trial, and cue  $\times$  trial interactions ( $F_{3,90} = 19.2$ ,  $P < 0.0001$ ;  $F_{5,150} = 28.8$ ,  $P < 0.0001$ ;  $F_{15,450} = 3.49$ ,  $P = 0.01$ , respectively). These effects reflected greater responding for A and B in responding across trials in the extinction probe tests, which was observed in both



**Figure 3 | Effect of orbitofrontal lesions on pavlovian conditioned responding after transreinforcer blocking in extinction probe tests.** Shown is the experimental timeline linked to data from the blocking probe tests. In the timeline and figures, A, B, X and Y are training cues; R1 and R2 are instrumental responses; and O1 and O2 are different flavoured sucrose pellet reinforcers. Responses per minute at the food cup are shown individually for

each of six unrewarded presentations of A, B, X and Y (small black (lesioned) and white (control) boxes) and also averaged over all six presentations (large grey (lesioned) and white (control) bars). Lesions had no effect on responding for A, B or X, but significantly diminished responding for Y. Error bars represent s.d.

controls and lesioned rats. ANOVA examining the effect of lesions on responding to the fully conditioned cues (A and B) yielded no significant main effects or any interactions ( $F < 1.2$ ,  $P > 0.29$  in all cases). In contrast, ANOVA examining the effect of lesions on responding to the blocked cues, X and Y, yielded a significant interaction between cue and lesion ( $F_{1,30} = 10.156$ ,  $P < 0.003$ ). Post-hoc tests showed that control rats responded significantly more for Y than for X ( $P < 0.001$ ) whereas lesioned rats did not ( $P > 0.34$ ). Thus, the OFC seems to be critical for both pavlovian and instrumental responses guided by information about the value of specific outcomes.

Here we have shown that conditioned reinforcers can drive behaviour either by activating representations of specific outcomes or by activating more general emotional or affective representations. These two types of information differ in critical ways: principally, the former reflects the animal's current desire for the outcome that the cue predicts, whereas the latter operates independently of that information. In normal settings, both affect and outcome information probably influence responding in concert; however, in theory, their influence on behaviour can be divergent. This was true in the current study for rats responding for the fully conditioned cue after devaluation of the outcome that this cue predicted. Notably, despite devaluation of the outcome, these rats continued to work for this cue. This imbalance is similar to the effects of conditioned reinforcers in humans, which are often independent of our desire for the outcomes they predict, as for example when one is lured into McDonald's despite a desire to avoid unhealthy fare. A similar but more severe imbalance is also prominent in neuropsychiatric disorders, including drug addiction and eating and anxiety disorders, in which cues come to control behaviour out of all proportion to the desirability of the predicted outcomes<sup>2,3</sup>.

The balance in how different types of associative information are mobilized by conditioned reinforcers may depend on the relative efficiency of processing in different neural circuits specialized for handling different types of associative information. Evidence presented here indicates that outcome-specific conditioned reinforcement is mediated by the OFC, whereas conditioned reinforcement mediated by general affective information is not. Although at least one previous report has implicated the OFC in conditioned reinforcement<sup>8</sup>, our results indicate that its role may be entirely subordinate to this region's well-documented involvement in signalling expected outcomes (see Supplementary Discussion for consideration of alternative interpretation involving attention). The OFC is activated by cues and during responding in a way that seems to reflect features of the specific outcomes that are to follow—particularly their value<sup>10–16</sup>—and damage to the OFC has been shown to impair changes in cue-evoked responding after devaluation<sup>17–19</sup>. Here we show that the OFC is also critical for instrumental responding when that responding is driven by cue-evoked representations of expected outcomes.

Of course the OFC is ultimately only one part of the brain circuit used by conditioned reinforcers to mobilize information. Areas such as the amygdala and the ventral striatum have also been implicated in responding for conditioned reinforcers<sup>20–27</sup>. In the same way that the OFC supports outcome-specific conditioned reinforcement, it seems likely that a subset of these other brain areas support conditioned reinforcement driven by cue-evoked representations of the general affect or emotion evoked by the cue<sup>28–30</sup>. An anatomical segregation of the brain circuits supporting conditioned reinforcement mediated by thoughts of specific outcomes versus more general affective information could help explain why disease states so often cause an imbalance in how we respond to conditioned reinforcers.

## METHODS SUMMARY

Thirty-two food-deprived male Long-Evans rats (Charles River Laboratories) served as subjects. Testing began after lesion/sham surgeries and was conducted using equipment from Coulbourn Instruments. Transreinforcer blocking consisted of pavlovian conditioning, compound conditioning, and probe tests. Rats received 12 days of conditioning in which two visual cues (a house light and a cue

light, designated A and B, counterbalanced) were paired 16 times with sucrose pellets (45 mg grape- and banana-flavoured sucrose pellets, designated O1 and O2, respectively, counterbalanced) (Research Diets).

After conditioning, all rats received 1 day of pre-exposure to two auditory cues (white noise and tone, designated X and Y, counterbalanced). The next day, the rats began 9 days of compound conditioning. Compound cue AX was paired with O1, which was the same flavour sucrose pellet associated with A. Compound BY was also paired with O1, which was a different flavour sucrose pellet than that associated with B. Each session consisted of eight presentations of each compound. The rats also received eight presentations of A–O1 and B–O2 reminder training. During compound conditioning, all rats received two probe test days consisting of six unrewarded presentations of A, B, X and Y.

Conditioned reinforcement was conducted during compound conditioning in the original training chambers with response levers on the right and left walls. One lever led to a 1-s presentation of a conditioned stimulus (either A or Y depending on group) on a fixed ratio two (FR2) schedule, and the second lever led to a 1-s presentation of X on an FR2 schedule. The lever–cue associations were counterbalanced for side in each group. Conditioned reinforcement testing was conducted once before, and again after, devaluation of O1. O1 was devalued by pairing it with illness induced by 0.3 M lithium chloride (5 mg kg<sup>−1</sup>, intraperitoneal injection).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 November 2007; accepted 8 April 2008.**

**Published online 18 June 2008.**

- Robinson, T. N., Borzekowski, D. L., Matheson, D. M. & Kraemer, H. C. Effects of fast food branding on young children's taste preferences. *Arch. Pediatr. Adolesc. Med.* **161**, 792–797 (2007).
- Shaham, Y., Shalev, U., Lu, L., de Wit, H. & Stewart, J. The reinstatement model of drug relapse: history, methodology and major findings. *Psychopharmacology (Berl.)* **168**, 3–20 (2003).
- O'Brien, C. P. Anticraving medications for relapse prevention: a possible new class of psychoactive medications. *Am. J. Psychiatry* **162**, 1423–1431 (2005).
- Murray, E. A., O'Doherty, J. P. & Schoenbaum, G. What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. *J. Neurosci.* **27**, 8166–8169 (2007).
- Rescorla, R. A. Learning about qualitatively different outcomes during a blocking procedure. *Anim. Learn. Behav.* **27**, 140–151 (1999).
- Ganesan, R. & Pearce, J. M. Effect of changing the unconditioned stimulus on appetitive blocking. *J. Exp. Psychol. Anim. Behav. Process.* **14**, 280–291 (1988).
- Weissenborn, R., Robbins, T. W. & Everitt, B. J. Effects of medial prefrontal or anterior cingulate cortex lesions on responding for cocaine under fixed-ratio and second-order schedules of reinforcement in rats. *Psychopharmacology (Berl.)* **134**, 242–257 (1997).
- Pears, A., Parkinson, J. A., Hopewell, L., Everitt, B. J. & Roberts, A. C. Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. *J. Neurosci.* **23**, 11189–11201 (2003).
- Parkinson, J. A., Roberts, A. C., Everitt, B. J. & Di Ciano, P. Acquisition of instrumental conditioned reinforcement is resistant to the devaluation of the unconditioned stimulus. *Q. J. Exp. Psychol.* **58**, 19–30 (2005).
- Padoa-Schioppa, C. & Assad, J. A. Neurons in orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
- Roesch, M. R., Taylor, A. R. & Schoenbaum, G. Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron* **51**, 509–520 (2006).
- Feierstein, C. E., Quirk, M. C., Uchida, N., Sosulski, D. L. & Mainen, Z. F. Representation of spatial goals in rat orbitofrontal cortex. *Neuron* **51**, 495–507 (2006).
- O'Doherty, J. P., Deichmann, R., Critchley, H. D. & Dolan, R. J. Neural responses during anticipation of a primary taste reward. *Neuron* **33**, 815–826 (2002).
- Gottfried, J. A., O'Doherty, J. & Dolan, R. J. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301**, 1104–1107 (2003).
- Tremblay, L. & Schultz, W. Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
- Roesch, M. R. & Olson, C. R. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* **304**, 307–310 (2004).
- Gallagher, M., McMahan, R. W. & Schoenbaum, G. Orbitofrontal cortex and representation of incentive value in associative learning. *J. Neurosci.* **19**, 6610–6614 (1999).
- Izquierdo, A., Suda, R. K. & Murray, E. A. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* **24**, 7540–7548 (2004).
- Pickens, C. L. et al. Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J. Neurosci.* **23**, 11078–11084 (2003).
- Setlow, B., Holland, P. C. & Gallagher, M. Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive Pavlovian second-order conditioned responses. *Behav. Neurosci.* **116**, 267–275 (2002).

21. Taylor, J. R. & Robbins, T. W. Enhanced behavioral control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. *Psychopharmacology (Berl.)* **84**, 405–412 (1984).
  22. Cousins, G. A. & Otto, T. Neural substrates of olfactory discrimination learning with auditory secondary reinforcement. I. Contributions of the basolateral amygdaloid complex and orbitofrontal cortex. *Integr. Physiol. Behav. Sci.* **38**, 272–294 (2003).
  23. Parkinson, J. A. *et al.* The role of the primate amygdala in conditioned reinforcement. *J. Neurosci.* **21**, 7770–7780 (2001).
  24. Setlow, B., Gallagher, M. & Holland, P. C. The basolateral complex of the amygdala is necessary for acquisition but not expression of CS motivational value in appetitive Pavlovian second-order conditioning. *Eur. J. Neurosci.* **15**, 1841–1853 (2002).
  25. Robledo, P., Robbins, T. W. & Everitt, B. J. Effects of excitotoxic lesions of the central amygdaloid nucleus on the potentiation of reward-related stimuli by intra-accumbens amphetamine. *Behav. Neurosci.* **110**, 981–990 (1996).
  26. Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W. & Everitt, B. J. Dissociation of effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by d-amphetamine. *J. Neurosci.* **19**, 2401–2411 (1999).
  27. Cador, M., Robbins, T. W. & Everitt, B. J. Involvement of the amygdala in stimulus–reward associations: interactions with the ventral striatum. *Neuroscience* **30**, 77–86 (1989).
  28. Hall, J., Parkinson, J. A., Connor, T. M., Dickinson, A. & Everitt, B. J. Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behavior. *Eur. J. Neurosci.* **13**, 1984–1992 (2001).
  29. Balleine, B. W. & Corbit, L. H. Double dissociation of nucleus accumbens core and shell on the general and outcome-specific forms of pavlovian-instrumental transfer. *Soc. Neurosci. Abstr.* **71** 16, (2005).
  30. Corbit, L. H. & Balleine, B. W. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. *J. Neurosci.* **25**, 962–970 (2005).
- Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).
- Acknowledgements** We are grateful to A. Delamater and P. Holland for their advice on this work. This work was supported by the National Institute on Drug Abuse (NIDA).
- Author Contributions** K.A.B. and G.S. conceived the experiments; K.A.B., D.N.M. and T.M.F. carried out the experiments; K.A.B. and G.S. analysed the data and co-wrote the manuscript with assistance from each of the other authors.
- Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.S. ([schoenbg@schoenbaumlab.org](mailto:schoenbg@schoenbaumlab.org)).



## METHODS

The order and timing of procedures is illustrated by the experimental timeline in Supplementary Fig. 1, which is reproduced from the main text.

**Subjects.** Thirty-two male Long-Evans rats (Charles River Laboratories) weighing between 275 and 300 g on arrival, were housed individually and placed on a 12-h light/dark schedule. All rats were given *ad libitum* access to food except during testing periods. During behavioural testing, rats were food-deprived to 85% of their baseline weight. All testing was conducted during the light period of their cycle.

**Surgical procedures.** Lesions were made in stereotaxic surgery using intracerebral infusions of *N*-methyl-D-aspartic acid (NMDA, Sigma) in saline vehicle. Infusions of 0.05–0.1  $\mu$ l of NMDA (12.5  $\mu$ g  $\mu$ l<sup>-1</sup>) were made at 4.0 mm anterior to the bregma, and at 2.2 and 3.7 mm lateral to the midline, at a depth of 3.8 mm ventral to the skull surface and 3.0 mm anterior to the bregma, at 3.2 mm lateral to the midline (0.05  $\mu$ l) and 4.2 mm lateral to the midline (0.1  $\mu$ l), and 5.2 mm ventral to the skull surface. Surgical controls received saline vehicle infusions. After a 1-week recovery period, all rats were placed on food restriction. Testing began 2 weeks after surgery (Supplementary Fig. 1).

**Apparatus.** Testing was conducted in eight standard sized behavioural boxes (12  $\times$  10  $\times$  12 inches) and other equipment modules purchased from Coulbourn Instruments. A recessed food cup was located in the centre of the right wall approximately 2 cm above the floor. The food cup was connected to a feeder mounted outside of the chamber to deliver 45 mg sucrose pellets (grape or banana, Research Diets). These pellets were equally preferred but discriminable. This was tested by giving food-deprived rats ( $n = 10$ ) access to both pellets before and after one of the two pellets was devalued by overfeeding (see Supplementary Fig. 3 for results). A house light and cue light were placed on the wall to the left or right of the food cup approximately 10 cm from the floor. Additionally, white noise and tone cues (75 dB, 4 kHz) could be delivered through speakers mounted in the centre of the wall. Data were collected by Graphic State behavioural software from Coulbourn Instruments.

**Transreinforcer blocking.** Transreinforcer blocking consisted of pavlovian conditioning, compound conditioning, and probe tests (Supplementary Fig. 1). Before any training, rats were reduced to 85% of their baseline weights, then they were shaped to the food cup during two training sessions in which two 45 mg sucrose pellets (grape and banana) were delivered to the food cup 16 times over the course of an hour. After these shaping sessions, all rats received 12 days of conditioning in which two visual cues (a house light and a cue light, designated A and B, counterbalanced) were paired with two distinctly flavoured yet equally

preferred sucrose pellets (45 mg grape and banana flavoured sucrose pellets, designated O1 and O2, counterbalanced; Research Diets). Sessions consisted of 16 presentations of each cue, with average inter-trial intervals of 2.5 min. During the 30-s presentation of each light cue, three food pellets were delivered with one food pellet being delivered every 8–10 s.

After conditioning, all rats received one day of pre-exposure to two auditory cues (white noise and tone, designated X and Y, counterbalanced). This pre-exposure consisted of one session in which each auditory cue was delivered six times for 30 s, with an average inter-trial interval of 2.5 min. The next day, the rats began 9 days of compound conditioning (Supplementary Fig. 1). In each session, compound cues, AX and BY, were presented for 30 s. Compound cue AX was paired with O1, which was the same flavour sucrose pellet associated with A. Compound BY was also paired with O1, which was a different flavour sucrose pellet than that associated with B. Each session consisted of eight presentations of each compound. The rats also received 8 presentations of A–O1 and B–O2 as reminder training; average inter-trial intervals were 2.5 min. During compound conditioning (Supplementary Fig. 1), all rats received two probe tests, each consisting of six unrewarded 30 s presentations of A, B, X and Y; average inter-trial intervals were 2.5 min.

**Conditioned reinforcement.** Conditioned reinforcement testing was conducted during compound conditioning, before and again after devaluation of the predicted outcome (Supplementary Fig. 1). Each episode of conditioned reinforcement testing consisted of two 30-min sessions and was conducted in the original training chambers with response levers inserted on the right and left walls. One lever led to a 1-s presentation of a conditioned stimulus (either A or Y depending on group) on a fixed ratio two (FR2) schedule, and the second lever led to a 1-s presentation of X on an FR2 schedule. The lever-cue associations were counterbalanced for side in each group.

**Devaluation.** The O1 food pellet was devalued by pairing it with illness induced by 0.3 M lithium chloride (5 mg kg<sup>-1</sup>, intraperitoneal injection). On days 1, 3 and 5 of this 6-day procedure, rats received access to the O1 pellet for 10 min in their home cage, followed by lithium chloride injection. On days 2, 4 and 6, rats received similar access to the O2 pellet with no injections. After all behavioural training, rats received a probe test in the home cage where they were given 10-min access to each pellet.

**Statistical analysis.** Data regarding food cup responding and bar pressing were acquired using Coulbourn GS2 software. Raw data were processed in Matlab to extract rates of responding. These data were analysed using Statistica.

# IGFBP-4 is an inhibitor of canonical Wnt signalling required for cardiogenesis

Weidong Zhu<sup>1\*</sup>, Ichiro Shiojima<sup>1\*</sup>, Yuzuru Ito<sup>2\*</sup>, Zhi Li<sup>1</sup>, Hiroyuki Ikeda<sup>1</sup>, Masashi Yoshida<sup>1</sup>, Atsuhiko T. Naito<sup>1</sup>, Jun-ichiro Nishi<sup>1</sup>, Hiroo Ueno<sup>3</sup>, Akihiro Umezawa<sup>4</sup>, Tohru Minamino<sup>1</sup>, Toshio Nagai<sup>1</sup>, Akira Kikuchi<sup>5</sup>, Makoto Asashima<sup>2,6,7</sup> & Issei Komuro<sup>1</sup>

**Insulin-like growth-factor-binding proteins (IGFBPs) bind to and modulate the actions of insulin-like growth factors (IGFs)<sup>1</sup>. Although some of the actions of IGFBPs have been reported to be independent of IGFs, the precise mechanisms of IGF-independent actions of IGFBPs are largely unknown<sup>1,2</sup>. Here we report a previously unknown function for IGFBP-4 as a cardiogenic growth factor. IGFBP-4 enhanced cardiomyocyte differentiation *in vitro*, and knockdown of *Igfbp4* attenuated cardiomyogenesis both *in vitro* and *in vivo*. The cardiogenic effect of IGFBP-4 was independent of its IGF-binding activity but was mediated by the inhibitory effect on canonical Wnt signalling. IGFBP-4 physically interacted with a Wnt receptor, Frizzled 8 (Frz8), and a Wnt co-receptor, low-density lipoprotein receptor-related protein 6 (LRP6), and inhibited the binding of Wnt3A to Frz8 and LRP6. Although IGF-independent, the cardiogenic effect of IGFBP-4 was attenuated by IGFs through IGFBP-4 sequestration. IGFBP-4 is therefore an inhibitor of the canonical Wnt signalling required for cardiogenesis and provides a molecular link between IGF signalling and Wnt signalling.**

The heart is the first organ to form during embryogenesis, and abnormalities in this process result in congenital heart diseases, the most common cause of birth defects in humans<sup>3</sup>. Molecules that mediate cardiogenesis are of particular interest because of their potential use for cardiac regeneration<sup>4,5</sup>. Previous studies have shown that soluble growth factors such as bone morphogenetic proteins (BMPs), fibroblast growth factors (FGFs), Wnts and Wnt inhibitors mediate the tissue interactions that are crucial for cardiomyocyte specification<sup>3,4</sup>. We proposed that there might be additional soluble factors that modulate cardiac development and/or cardiomyocyte differentiation.

P19CL6 cells differentiate into cardiomyocytes with high efficiency in the presence of 1% dimethylsulphoxide (DMSO)<sup>6</sup>. We cultured P19CL6 cells with culture media conditioned by various cell types in the absence of DMSO, and screened the cardiogenic activity of the conditioned media. The extent of cardiomyocyte differentiation was assessed by the immunostaining with MF20 monoclonal antibody that recognizes sarcomeric myosin heavy chain (MHC). Among the several cell types tested, culture media conditioned by a murine stromal cell line OP9 induced cardiomyocyte differentiation of P19CL6 cells without DMSO treatment (Fig. 1a, left and middle panels). Increased MF20-positive area was accompanied by the induction of cardiac marker genes such as  $\alpha$ MHC, *Nkx2.5* and *GATA-4*, and by the increased protein levels of cardiac troponin T (cTnT) (Fig. 1a,

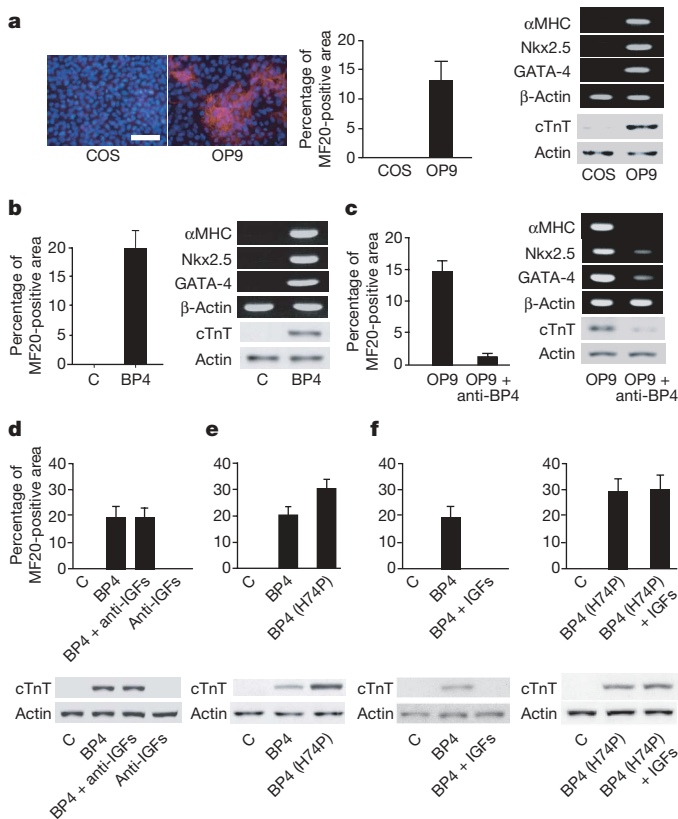
right panel). In contrast, culture media conditioned by COS7 cells, mouse embryonic fibroblasts, NIH3T3 cells, HeLa cells, END2 cells (visceral endoderm-like cells), neonatal rat cardiomyocytes and neonatal rat cardiac fibroblasts did not induce cardiomyocyte differentiation of P19CL6 cells in the absence of DMSO (Fig. 1a and data not shown). From these observations, we postulated that OP9 cells secrete one or more cardiogenic growth factors.

To identify an OP9-derived cardiogenic factor, complementary DNA clones isolated by a signal sequence trap method from an OP9 cell cDNA library<sup>7</sup> were tested for their cardiogenic activities by transient transfection. When available, recombinant proteins were also used to confirm the results. Among candidate factors tested, IGFBP-4 induced cardiomyocyte differentiation of P19CL6 cells, as demonstrated by the increase in MF20-positive area and the induction of cardiac markers (Fig. 1b). We also cultured P19CL6 cells with OP9-conditioned media pretreated with an anti-IGFBP-4 neutralizing antibody. The application of an anti-IGFBP-4 neutralizing antibody attenuated the efficiency of cardiomyocyte differentiation induced by OP9-conditioned media (Fig. 1c). These findings strongly suggest that IGFBP-4 is a cardiogenic factor secreted from OP9 cells.

Because IGFBPs have been characterized as molecules that bind to and modulate the actions of IGFs, we tested whether IGFBP-4 promotes cardiogenesis by either enhancing or inhibiting the actions of IGFs. We first treated P19CL6 cells with a combination of anti-IGF-I and IGF-II-neutralizing antibodies or a neutralizing antibody against type-I IGF receptor. If IGFBP-4 induces cardiomyocyte differentiation by inhibiting IGF signalling, treatment with these antibodies should induce cardiomyocyte differentiation and/or enhance the cardiogenic effects of IGFBP-4. In contrast, if IGFBP-4 promotes cardiogenesis by enhancing IGF signalling, treatment with these antibodies should attenuate IGFBP-4-mediated cardiogenesis. However, treatment with these antibodies did not affect the efficiency of IGFBP-4-induced cardiomyocyte differentiation (Fig. 1d and data not shown). Treatment of P19CL6 cells with IGF-I and IGF-II also did not induce cardiomyocyte differentiation (data not shown). Furthermore, treatment with an IGFBP-4 mutant (IGFBP-4-H74P; His 74 replaced by Pro)<sup>8</sup> that is unable to bind IGFs induced cardiomyocyte differentiation of P19CL6 cells even more efficiently than wild-type IGFBP-4 (Fig. 1e). This is presumably due to the sequestration of wild-type IGFBP-4 but not mutant IGFBP-4-H74P by endogenous IGFs. In agreement with this idea, exogenous IGFs attenuated wild-type IGFBP-4-induced but not IGFBP-4-H74P-induced cardiogenesis (Fig. 1f). Taken together, these observations indicate

<sup>1</sup>Department of Cardiovascular Science and Medicine, Chiba University Graduate School of Medicine, Chiba 260-8670, Japan. <sup>2</sup>ICORP Organ Regeneration Project, Japan Science and Technology Agency (JST), Tokyo 153-8902, Japan. <sup>3</sup>Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>4</sup>Department of Reproductive Biology, National Institute for Child Health and Development, Tokyo 157-8535, Japan. <sup>5</sup>Department of Biochemistry, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima 734-8551, Japan. <sup>6</sup>Department of Life Sciences (Biology), Graduate School of Arts and Science, The University of Tokyo, Tokyo 153-8902, Japan. <sup>7</sup>National Institute of Advanced Industrial Sciences and Technology (AIST), Ibaraki 305-8562, Japan.

\*These authors contributed equally to this work.



**Figure 1 | IGFBP-4 promotes cardiomyocyte differentiation in an IGF-independent manner.** **a**, Culture media conditioned by OP9 cells but not by COS7 cells induced cardiomyocyte differentiation of P19CL6 cells as assessed by MF20-positive area, cardiac marker-gene expression and cTnT protein expression. Scale bar, 100  $\mu$ m. Error bars show s.d. **b**, Treatment with IGFBP-4 ( $1 \mu\text{g ml}^{-1}$ ) induced cardiomyocyte differentiation of P19CL6 cells in the absence of DMSO. Error bars show s.d. **c**, Treatment with a neutralizing antibody against IGFBP-4 (anti-BP4;  $40 \mu\text{g ml}^{-1}$ ) attenuated cardiomyocyte differentiation of P19CL6 cells induced by OP9-conditioned media. Error bars show s.d. **d**, Treatment with neutralizing antibodies against IGF-I and IGF-II (anti-IGFs;  $5 \mu\text{g ml}^{-1}$  each) had no effect on IGFBP-4-induced cardiomyocyte differentiation of P19CL6 cells. Error bars show s.d. **e**, Mutant IGFBP-4 (BP4(H74P)) that is incapable of binding to IGFs retained cardiomyogenic activity. Error bars show s.d. **f**, IGFs ( $100 \text{ ng ml}^{-1}$  each) attenuated wild-type IGFBP-4-induced but not mutant IGFBP-4-H74P-induced cardiomyocyte differentiation of P19CL6 cells. Error bars show s.d.

that IGFBP-4 induces cardiomyocyte differentiation in an IGF-independent fashion.

To explore further the mechanisms by which IGFBP-4 induces cardiomyogenesis, we tested the hypothesis that IGFBP-4 might modulate the signals activated by other secreted factors implicated in cardiogenesis. It has been shown that canonical Wnt signalling is crucial in cardiomyocyte differentiation<sup>3,4</sup>. In P19CL6 cells, Wnt3A treatment activated  $\beta$ -catenin-dependent transcription of the TOPFLASH reporter gene, and this activation was attenuated by IGFBP-4 (Fig. 2a). Wnt/ $\beta$ -catenin signalling is transduced by the cell-surface receptor complex consisting of Frizzled and low-density-lipoprotein receptor (LDLR)-related protein 5/6 (LRP5/6)<sup>9</sup> and IGFBP-4 attenuated TOPFLASH activity enhanced by the expression of LRP6 or Frizzled 8 (Frz8) (Fig. 2a). As a control, IGFBP-4 did not alter BMP-mediated activation of a BMP-responsive reporter BRE-luc (Supplementary Fig. 1b). These findings suggest that IGFBP-4 is a specific inhibitor of the canonical Wnt pathway. To examine this possibility *in vivo*, we performed axis duplication assays in *Xenopus* embryos. Injection of *Xwnt8* or *Lrp6* mRNA caused secondary axis formation, and injection of *Xenopus* IGFBP-4 (*XIGFBP-4*) mRNA alone had minimal effects on axis

formation. However, *Xwnt8*-induced or LRP6-induced secondary axis formation was efficiently blocked by coexpression of *XIGFBP-4* (Fig. 2b, c), indicating that IGFBP-4 inhibits canonical Wnt signalling *in vivo*. To explore the mechanisms of Wnt inhibition by IGFBP-4, *Xenopus* animal cap assays and TOPFLASH reporter gene assays were performed. In animal cap assays, IGFBP-4 inhibited LRP6-induced but not  $\beta$ -catenin-induced Wnt-target gene expression (Supplementary Fig. 1c). Similarly, IGFBP-4 attenuated Wnt3A-induced or LRP6-induced TOPFLASH activity but did not alter Dishevelled-1 (Dvl-1)-induced, LiCl-induced or  $\beta$ -catenin-induced TOPFLASH activity (Supplementary Fig. 1d, e). These findings suggest that IGFBP-4 inhibits canonical Wnt signalling at the level of cell-surface receptors. To examine whether IGFBP-4 antagonizes Wnt signalling via direct physical interaction with LRP5/6 or Frizzled, we produced conditioned media containing the Myc-tagged extracellular portion of LRP6 (LRP6N-Myc), the Myc-tagged cysteine-rich domain (CRD) of Frz8 (Frz8CRD-Myc), and V5-tagged IGFBP-4 (IGFBP-4-V5). Immunoprecipitation (IP)/western blot experiments revealed that IGFBP-4 interacted with LRP6N (Fig. 2d) and Frz8CRD (Fig. 2e). A liquid-phase binding assay with <sup>125</sup>I-labelled IGFBP-4 and conditioned media containing LRP6N-Myc or Frz8CRD-Myc demonstrated that the interaction between IGFBP-4 and LRP6N or Frz8CRD was specific and saturable (Fig. 2f, g). A Scatchard plot analysis revealed two binding sites with different binding affinities for LRP6N (Fig. 2f, inset) and a single binding site for Frz8CRD (Fig. 2g, inset). A similar binding assay with <sup>125</sup>I-labelled Wnt3A demonstrated that IGFBP-4 inhibited Wnt3A binding to LRP6N (Fig. 2h) and Frz8CRD (Fig. 2i), and a Lineweaver–Burk plot revealed that IGFBP-4 was a competitive inhibitor of the binding of Wnt3A to Frz8CDR (Supplementary Fig. 2a). IP/western blot analyses with various deletion mutants of LRP6 and IGFBP-4 revealed that IGFBP-4 interacted with multiple domains of LRP6 and that the carboxy-terminal thyroglobulin domain of IGFBP-4 was required for IGFBP-4 binding to LRP6 or Frz8CRD (Supplementary Fig. 2b–f). It has been shown that inhibition of canonical Wnt signalling promotes cardiomyocyte differentiation in embryonic stem (ES) cells and in chick, *Xenopus* and zebrafish embryos<sup>4,10,11</sup>. These results therefore collectively suggest that IGFBP-4 promotes cardiogenesis by antagonizing the Wnt/ $\beta$ -catenin pathway through direct interactions with Frizzled and LRP5/6.

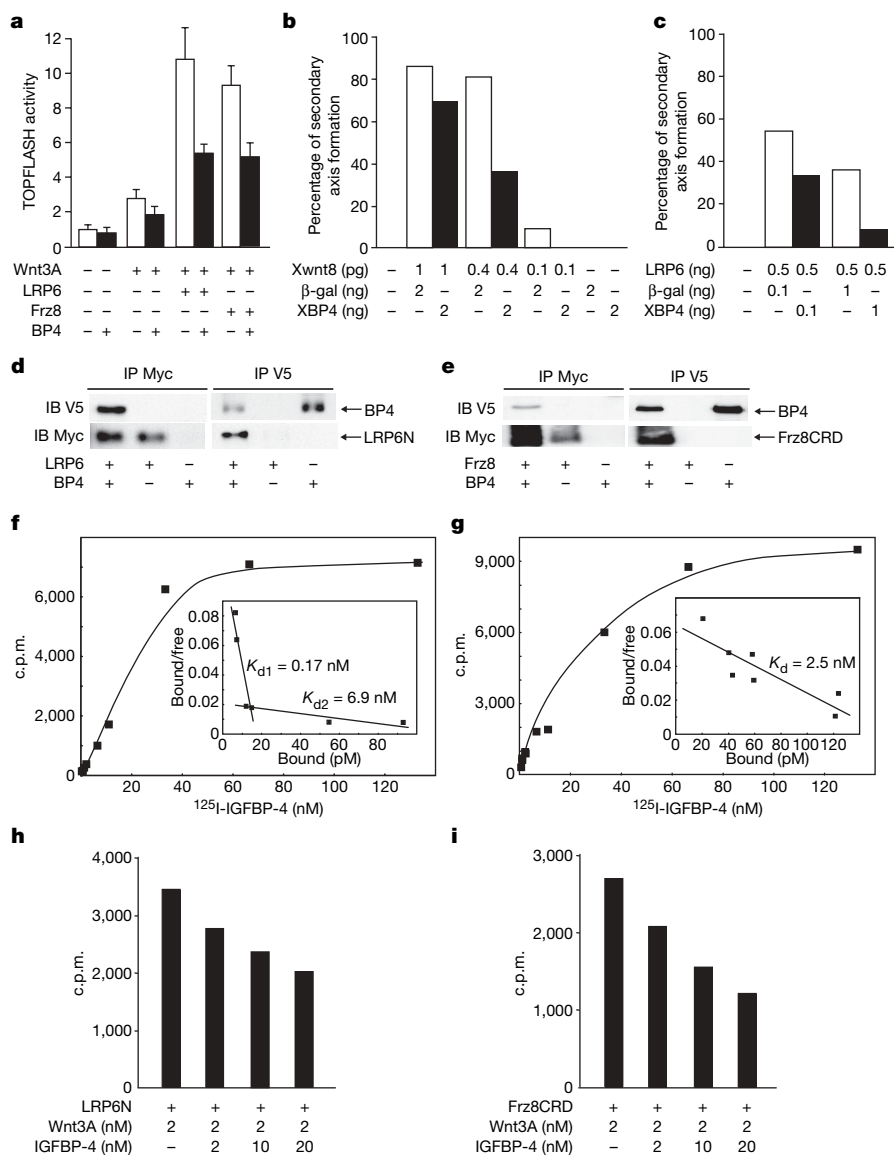
Next we investigated the role of endogenous IGFBP-4 in P19CL6 cell differentiation into cardiomyocytes. Reverse transcriptase-mediated polymerase chain reaction (RT-PCR) analysis revealed that the expression of *Igfbp4* was upregulated during DMSO-induced P19CL6 cell differentiation (Fig. 3a). Expression of *Igfbp3* and *Igfbp5* was also upregulated in the early and the late phases of differentiation, respectively. Expression of *Igfbp2* was not altered, and that of *Igfbp1* or *Igfbp6* was not detected. When IGFBP-4 was knocked down by two different small interfering RNA (siRNA) constructs, DMSO-induced cardiomyocyte differentiation was inhibited in both cases (Fig. 3b). In contrast, knockdown of *Igfbp3* or *Igfbp5* did not inhibit DMSO-induced cardiomyocyte differentiation (Fig. 3b, right panel). Treatment with an anti-IGFBP-4 neutralizing antibody also blocked DMSO-induced cardiomyocyte differentiation (Fig. 3c). Secretion of endogenous IGFBP-4 is therefore required for the differentiation of P19CL6 cells into cardiomyocytes. Immunostaining for IGFBP-4 revealed that cardiac myocytes were surrounded by the IGFBP-4-positive cells, suggesting that a paracrine effect of IGFBP-4 on cardiomyocyte differentiation is predominant (Fig. 3d). Essentially the same results were obtained in ES cells (Supplementary Fig. 3d–g). To investigate whether IGFBP-4 promotes the differentiation of P19CL6 cells into cardiomyocytes by the inhibition of the canonical Wnt pathway, we expressed dominant-negative LRP6 (LRP6N) in P19CL6 cells. Expression of LRP6N enhanced cardiomyocyte differentiation of P19CL6 cells and reversed the inhibitory effect of *Igfbp4*



knockdown on cardiomyogenesis (Fig. 3e). These observations suggest that endogenous IGFBP-4 is required for cardiomyocyte differentiation of P19CL6 cells and ES cells, and that the cardiogenic effect of IGFBP-4 is mediated by its inhibitory effect on Wnt/ $\beta$ -catenin signalling.

The role of endogenous IGFBP-4 in cardiac development *in vivo* was also examined with *Xenopus* embryos. Whole-mount *in situ* hybridization analysis revealed that strong expression of *XIGFBP-4* was detected at stage 38 in the anterior part of the liver adjacent to the heart (Fig. 4a). Knockdown of *XIGFBP-4* by two different morpholino (MO) constructs resulted in cardiac defects, with more than 70% of the embryos having a small heart or no heart (Fig. 4b). The specificity of MO was confirmed by the observation that simultaneous injection of MO-resistant *XIGFBP-4* cDNA rescued the MO-induced cardiac defects (Fig. 4b, Supplementary Fig. 4c). Coexpression of IGF-binding-defective *XIGFBP-4* mutant (*XIGFBP-4*-H74P) or

dominant-negative LRP6 (LRP6N) also rescued the cardiac defects induced by *XIGFBP-4* knockdown (Fig. 4b), whereas overexpression of Xwnt8 in the heart-forming region resulted in cardiac defects similar to those induced by *XIGFBP-4* knockdown (Supplementary Fig. 4d–f), supporting the notion that the cardiogenic effect of IGFBP-4 is independent of IGFs but is mediated by inhibition of the Wnt/ $\beta$ -catenin pathway. The temporal profile of cardiac defects induced by *XIGFBP-4* knockdown was also examined by *in situ* hybridization with *cardiac troponin I* (*cTnI*) (Fig. 4c). At stage 34, morphology of the heart was comparable between control embryos and MO-injected embryos. However, at stage 38, when *XIGFBP-4* starts to be expressed in the anterior part of the liver, the expression of *cTnI* was markedly attenuated in MO-injected embryos; expression of *cTnI* was diminished and no heart-like structure was observed at stage 42. Thus, the heart is initially formed but its subsequent growth is perturbed in the absence of *XIGFBP-4*, suggesting that IGFBP-4

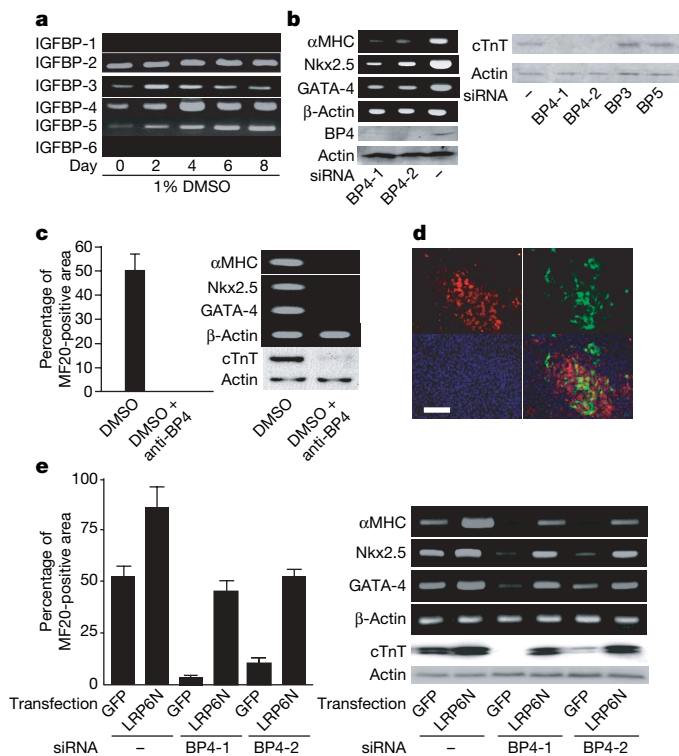


**Figure 2** | IGFBP-4 inhibits Wnt/ $\beta$ -catenin signalling through direct interactions with Wnt receptors. **a**, IGFBP-4 attenuated  $\beta$ -catenin-dependent transcription in P19CL6 cells. P19CL6 cells were transfected with TOPFLASH reporter gene and expression vectors for LRP6 or Frz8, and then treated with Wnt3A or Wnt3A plus IGFBP-4; luciferase activities were then measured. Error bars show s.d. **b**, *XIGFBP-4* (XBP4) inhibited Xwnt8-induced secondary-axis formation in *Xenopus* embryos ( $n = 20$  for each group). **c**, IGFBP-4 inhibited LRP6-induced secondary-axis formation in *Xenopus* embryos ( $n = 30$  for each group). **d**, **e**, IGFBP-4 interacted directly

with LRP6N (**d**) and Frz8CRD (**e**). IB, immunoblotting; IP, immunoprecipitation. **f**, A binding assay between  $^{125}$ I-labelled IGFBP-4 and LRP6N. The inset is a Scatchard plot showing two binding sites with different binding affinities. **g**, A binding assay between  $^{125}$ I-labelled IGFBP-4 and Frz8CRD. The inset is a Scatchard plot showing a single binding site. **h**, **i**, IGFBP-4 inhibited Wnt3A binding to LRP6N (**h**) or Frz8CRD (**i**).  $^{125}$ I-labelled Wnt3A binding to LRP6N or Frz8CRD was assessed in the presence of increasing amounts of IGFBP-4.

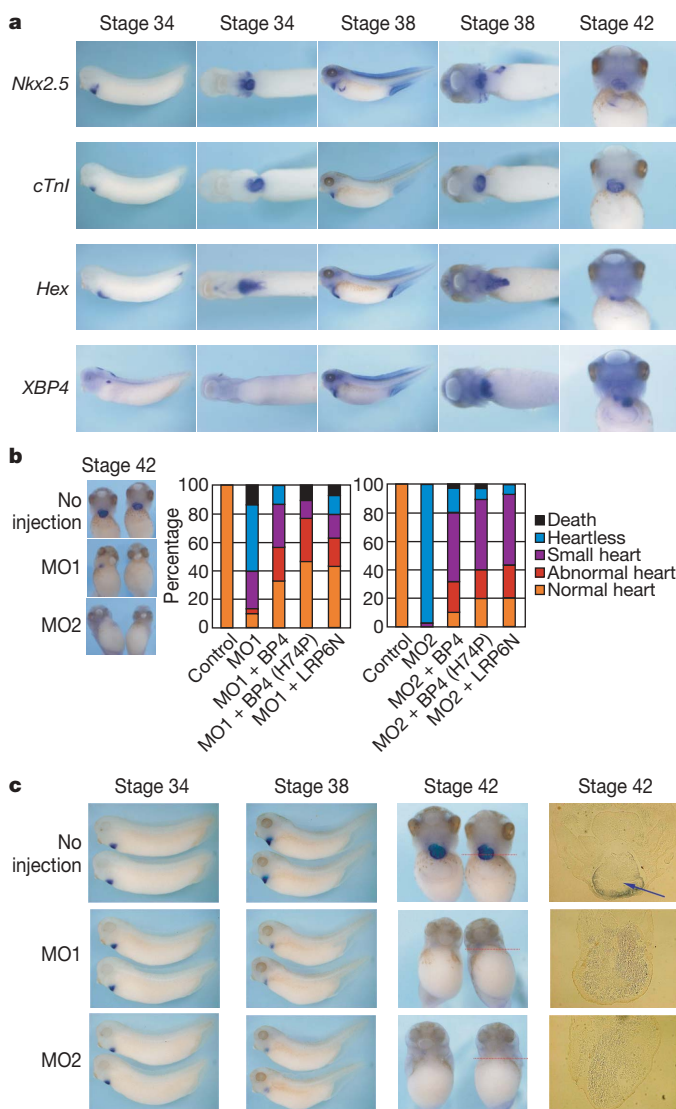
promotes cardiogenesis by maintaining the proliferation and/or survival of embryonic cardiomyocytes.

It has been shown that canonical Wnt signals inhibit cardiogenesis in chick and frog embryos, and that Wnt antagonists such as Dkk1 and Crescent secreted from the anterior endoderm or the organizer region counteract the Wnt-mediated inhibitory signals and induce cardiogenesis in the anterior lateral mesoderm<sup>4</sup>. However, IGFBP-4-mediated Wnt inhibition is required at later stages of development, when the heart is already formed at the ventral portion and starts to grow and remodel to maintain embryonic circulation. It has been shown that Wnt/ $\beta$ -catenin signalling has time-dependent effects on cardiogenesis in ES cells: canonical Wnt signalling in the early phase of ES-cell differentiation promotes cardiomyogenesis, whereas it inhibits cardiomyocyte differentiation in the late phase<sup>10–12</sup>. In agreement with this notion, IGFBP-4 promoted cardiomyocyte differentiation of ES cells only when IGFBP-4 was applied in the late phase after embryoid body formation (Supplementary Fig. 3a–c). Similar



**Figure 3 | IGFBP-4 is required for the differentiation of P19CL6 cells into cardiomyocytes.** **a**, Expression analysis of IGFBP family members by RT-PCR during DMSO-induced cardiomyocyte differentiation of P19CL6 cells (from day 0 to day 8). **b**, Left: knockdown of *Igfbp4* in P19CL6 cells attenuated cardiac marker expression in response to treatment with DMSO. BP4-1 and BP4-2 represent two different siRNAs for IGFBP-4. Right: knockdown of *Igfbp3* or *Igfbp5* had no effect on cTnT expression in response to DMSO treatment. **c**, Treatment with a neutralizing antibody against IGFBP-4 (anti-BP4; 40  $\mu\text{g ml}^{-1}$ ) attenuated DMSO-induced cardiomyocyte differentiation of P19CL6 cells. Error bars show s.d. **d**, IGFBP-4 immunostaining during DMSO-induced differentiation of P19CL6 cells stably transfected with  $\alpha\text{MHC}$ -green fluorescent protein (GFP) reporter gene. Top left, IGFBP-4 staining (red); top right, GFP expression representing differentiated cardiomyocytes; bottom left, nuclear staining with DAPI (4',6-diamidino-2-phenylindole); bottom right, a merged picture. Scale bar, 100  $\mu\text{m}$ . **e**, Attenuated cardiomyocyte differentiation of P19CL6 cells by *Igfbp4* knockdown was rescued by inhibiting Wnt/ $\beta$ -catenin signalling. Control and *Igfbp4*-knocked-down P19CL6 cells were transfected with an expression vector for GFP or LRP6N (a dominant-negative form of LRP6) and induced to differentiate into cardiomyocytes by treatment with DMSO. LRP6N overexpression rescued the attenuated cardiomyocyte differentiation induced by *Igfbp4* knockdown as assessed by MF20-positive area (left panel), cardiac marker-gene expression and cTnT protein expression (right panel). Error bars show s.d.

time-dependent effects of Wnt/ $\beta$ -catenin signalling on cardiogenesis has been shown in zebrafish embryos<sup>11</sup>. Moreover, several recent reports suggest that Wnt/ $\beta$ -catenin signalling is a positive regulator of cardiac progenitor-cell proliferation in the secondary heart field<sup>13</sup>. It therefore seems that canonical Wnt signalling has divergent effects on cardiogenesis at multiple stages of development: first, canonical Wnt signalling promotes cardiogenesis at the time of gastrulation or mesoderm specification; second, it inhibits cardiogenesis at the time when cardiac mesoderm is specified in the anterior lateral mesoderm; third, it promotes the expansion of cardiac progenitors in the secondary heart field; and fourth, it inhibits cardiogenesis at later stages when the embryonic heart is growing. It is interesting to note that IGFBP-4 is expressed predominantly in the liver. Mouse IGFBP-4 is



**Figure 4 | IGFBP-4 is required for the maturation of the heart in *Xenopus* embryos.** **a**, *In situ* hybridization analysis of *Nkx2.5* (an early cardiac marker), *cTnI* (a mature cardiac marker), *Hex* (a liver marker), and *XIGFBP-4* (*XBP4*) mRNA expression at stages 34, 38 and 42. **b**, Knockdown of *XIGFBP-4* by two different morpholinos (MO1 and MO2) resulted in severe cardiac defects as assessed by *cTnI* *in situ* hybridization at stage 42 (left). These cardiac defects were rescued by simultaneous injection of MO-resistant wild-type *XIGFBP-4*, mutant *XIGFBP-4*-H74P (BP4(H74P)) and LRP6N ( $n = 30$  for each group). **c**, Temporal profile of cardiac defects induced by *XIGFBP-4* knockdown. Morphology of the heart as assessed by *cTnI* *in situ* hybridization was almost normal at stage 34 but was severely perturbed at stages 38 and 42. The right column shows sections of control and MO-injected embryos. The arrow indicates the heart in control embryos. No heart-like structure was observed in MO-injected embryos.

also strongly expressed in the tissues adjacent to the heart such as pharyngeal arches and liver bud at embryonic day (E)9.5 (Supplementary Fig. 3h). These observations and the results of IGFBP-4 immunostaining in P19CL6 cells and ES cells suggest that IGFBP-4 promotes cardiogenesis in a paracrine fashion. Together with a previous report showing that cardiac mesoderm secretes FGFs and induces liver progenitors in the ventral endoderm<sup>14</sup>, these observations suggest that there exist reciprocal paracrine signals between the heart and the liver that coordinately promote the development of each other.

IGFBPs are composed of six members, IGFBP-1 to IGFBP-6. Reporter gene assays and  $\beta$ -catenin stabilization assays revealed that IGFBP-4 was the most potent canonical Wnt inhibitor and that IGFBP-1, IGFBP-2 and IGFBP-6 also showed modest activity in Wnt inhibition, whereas IGFBP-3 and IGFBP-5 had no such activity (Supplementary Fig. 5a–c). In agreement with this, IP/western blot analyses demonstrated that IGFBP-1, IGFBP-2, IGFBP-4 and IGFBP-6 but not IGFBP-3 or IGFBP-5 interacted with LRP6 or Frz8CRD (Supplementary Fig. 5d, e). Thus, the lack of cardiac phenotypes in IGFBP-4-null mice or IGFBP-3/IGFBP-4/IGFBP-5 triple knockout mice<sup>15</sup> may be due to genetic redundancies between IGFBP-4 and other IGFBPs such as IGFBP-1, IGFBP-2 and/or IGFBP-6.

The identification of IGFBP-4 as an inhibitor of Wnt/ $\beta$ -catenin signalling may also have some implications for cancer biology<sup>16</sup>. It was shown that treatment with IGFBP-4 reduces cell proliferation in some cancer cell lines *in vitro*, and that overexpression of IGFBP-4 attenuates the growth of prostate cancer *in vivo*. Decreased serum levels of IGFBP-4 are associated with the risk of breast cancer. Because the activation of Wnt signalling is implicated in several forms of malignant tumours<sup>17,18</sup>, it is possible that the inhibitory effect of IGFBP-4 on cell proliferation is mediated in part by the inhibition of canonical Wnt signalling.

## METHODS SUMMARY

**Cell culture.** P19CL6 cells and ES cells were cultured and induced to differentiate into cardiomyocytes essentially as described<sup>6,10</sup>. P19CL6 cells (2,000 cells per 35-mm dish) were treated with various conditioned media for screening of their cardiogenic activities. For siRNA-mediated knockdown, pSIREN-RetroQ vectors (Clontech) ligated with double-stranded oligonucleotides were transfected into P19CL6 cells or ES cells, and puromycin-resistant clones were selected.

**IP/western blot analyses and binding assays.** Conditioned media for IP/western blot analyses were produced by using 293 cells. Binding reactions were performed overnight at 4 °C. <sup>125</sup>I-labelling of IGFBP-4 and Wnt3A was performed with IODO-BEADS Iodination Reagent (Pierce). A liquid-phase binding assay was performed essentially as described<sup>19</sup>.

**Xenopus experiments.** Axis duplication assays, animal cap assays, and *in situ* hybridization analyses in *Xenopus* were performed essentially as described<sup>20</sup>. Electroporation of mRNA was performed at stage 28 essentially as described<sup>21</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 August 2007; accepted 24 April 2008.

Published online 4 June 2008.

1. Firth, S. M. & Baxter, R. C. Cellular actions of the insulin-like growth factor binding proteins. *Endocr. Rev.* **23**, 824–854 (2002).
2. Mohan, S. & Baylink, D. J. IGF-binding proteins are multifunctional and act via IGF-dependent and -independent mechanisms. *J. Endocrinol.* **175**, 19–31 (2002).

3. Olson, E. N. & Schneider, M. D. Sizing up the heart: development redux in disease. *Genes Dev.* **17**, 1937–1956 (2003).
4. Foley, A. & Mercola, M. Heart induction: embryology to cardiomyocyte regeneration. *Trends Cardiovasc. Med.* **14**, 121–125 (2004).
5. Leri, A., Kajstura, J. & Anversa, P. Cardiac stem cells and mechanisms of myocardial regeneration. *Physiol. Rev.* **85**, 1373–1416 (2005).
6. Monzen, K. *et al.* Bone morphogenetic proteins induce cardiomyocyte differentiation through the mitogen-activated protein kinase kinase TAK1 and cardiac transcription factors Csx/Nkx-2.5 and GATA-4. *Mol. Cell. Biol.* **19**, 7096–7105 (1999).
7. Ueno, H. *et al.* A stromal cell-derived membrane protein that supports hematopoietic stem cells. *Nature Immunol.* **4**, 457–463 (2003).
8. Qin, X., Strong, D. D., Baylink, D. J. & Mohan, S. Structure–function analysis of the human insulin-like growth factor binding protein-4. *J. Biol. Chem.* **273**, 23509–23516 (1998).
9. Moon, R. T., Kohn, A. D., De Ferrari, G. V. & Kaykas, A. WNT and  $\beta$ -catenin signalling: diseases and therapies. *Nature Rev. Genet.* **5**, 691–701 (2004).
10. Naito, A. T. *et al.* Developmental stage-specific biphasic roles of Wnt/ $\beta$ -catenin signaling in cardiomyogenesis and hematopoiesis. *Proc. Natl Acad. Sci. USA* **103**, 19812–19817 (2006).
11. Ueno, S. *et al.* Biphasic role for Wnt/ $\beta$ -catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 9685–9690 (2007).
12. Liu, Y. *et al.* Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 3859–3864 (2007).
13. Cohen, E. D., Tian, Y. & Morrisey, E. E. Wnt signaling: an essential regulator of cardiovascular differentiation, morphogenesis and progenitor self-renewal. *Development* **135**, 789–798 (2008).
14. Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* **284**, 1998–2003 (1999).
15. Ning, Y. *et al.* Diminished growth and enhanced glucose metabolism in triple knockout mice containing mutations of insulin-like growth factor binding protein-3, -4, and -5. *Mol. Endocrinol.* **20**, 2173–2186 (2006).
16. Durai, R. *et al.* Biology of insulin-like growth factor binding protein-4 and its role in cancer. *Int. J. Oncol.* **28**, 1317–1325 (2006).
17. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
18. Clevers, H. Wnt/ $\beta$ -catenin signaling in development and disease. *Cell* **127**, 469–480 (2006).
19. Semenov, M. V. *et al.* Head inducer Dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Curr. Biol.* **11**, 951–961 (2001).
20. Kobayashi, H. *et al.* Novel Daple-like protein positively regulates both the Wnt/ $\beta$ -catenin pathway and the Wnt/JNK pathway in *Xenopus*. *Mech. Dev.* **122**, 1138–1153 (2005).
21. Sasagawa, S., Takabatake, T., Takabatake, Y., Muramatsu, T. & Takeshima, K. Improved mRNA electroporation method for *Xenopus* neurula embryos. *Genesis* **33**, 81–85 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Fujita, R. Kobayashi and Y. Ishiyama for technical support; T. Yamauchi and K. Ueki for advice on binding assays; and Y. Onuma and S. Takahashi for advice on *Xenopus* electroporation. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Ministry of Health, Labour, and Welfare, and the New Energy and Industrial Technology Development Organization (NEDO).

**Author Contributions** W.Z., I.S. and Y.I. contributed equally to this work. I.K. designed and supervised the research. W.Z., I.S., Y.I., Z.L., H.L., M.Y. and A.T.N. performed experiments. J.N., H.U., A.U., T.M., T.N., A.K. and M.A. contributed new reagents and/or analytical tools. W.Z., I.S., Y.I., A.K. and I.K. analysed data. W.Z., I.S., Y.I. and I.K. prepared the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to I.K. ([komuro-ty@umin.ac.jp](mailto:komuro-ty@umin.ac.jp)).



## METHODS

**Plasmids and reagents.** cDNA clones encoding mouse IGFBPs and *Xenopus* IGFBP-4 were purchased from Open Biosystems. XIGFBP-4-H74P mutant was generated with a QuickChange Site-Directed Mutagenesis kit (Stratagene). His-tagged human wild-type IGFBP-4 and mutant IGFBP-4-H74P (vectors provided by X. Qin)<sup>8</sup> were produced and purified with HisTrap HP Kit (Amersham). Full-length Frz8, Frz8CRD and LRP6N were provided by X. He<sup>22,23</sup>. Full-length LRP6, membrane-bound forms of LRP6 deletion mutants, and Dkk1 were from C. Niehrs<sup>24</sup>. pXwnt8 and pCSKA-Xwnt8 were from J. Christian<sup>25</sup>. pCS2- $\beta$ -catenin was from D. Kimelman<sup>26</sup>.  $\alpha$ MHC-GFP was from B. Fleischmann<sup>27</sup>. BRE-luc was from P. ten Dijke<sup>28</sup>. pCGN-Dvl-1 was described previously<sup>29</sup>. Soluble forms of LRP6 deletion mutants and probes for *in situ* hybridization analysis (Nkx2.5, cTnI and Hex) were generated by PCR. IGFBP-4, Wnt3A, IGF-I, IGF-II and BMP2 were from R&D. Neutralizing antibodies were from R&D (anti-IGFBP-4), Sigma (anti-IGF-I and anti-IGF-II), and Oncogene (anti-type-I IGF receptor). The antibodies used for immunoprecipitation, western blotting and immunostaining were from Invitrogen (anti-Myc, anti-V5), Santa Cruz (anti-cTnT, anti-IGFBP-4, anti-topoisomerase I (TOPO-I)), Sigma (anti- $\beta$ -actin, anti- $\beta$ -catenin, anti-FLAG (M2)) and Developmental Studies Hybridoma Bank (anti-sarcomeric myosin heavy chain (MF20)).

**Cell culture experiments.** P19CL6 cells and ES cells were cultured and induced to differentiate into cardiomyocytes essentially as described<sup>6,10</sup>. P19CL6 cells (2,000 cells per 35-mm dish) were treated with various conditioned media for screening of their cardiogenic activities. P19CL6 cells or ES cells stably transfected with  $\alpha$ MHC promoter driven-GFP were generated by transfection of  $\alpha$ MHC-GFP plasmid into P19CL6 cells or h7 ES cells followed by G418 selection. Luciferase reporter gene assays, western blot analyses, immunostaining and RT-PCR were performed as described<sup>10</sup>. Reporter gene assays were repeated at least three times. PCR primers and PCR conditions are listed in Supplementary Table 1. For siRNA-mediated knockdown, siRNAs were expressed with pSIREN-RetroQ vector (Clontech). Oligonucleotide sequences used are listed in Supplementary Table 2. pSIREN-RetroQ vectors ligated with double-stranded oligonucleotides were transfected into P19CL6 cells or ES cells, and puromycin-resistant clones were isolated and expanded. For  $\beta$ -catenin stabilization assays, nuclear extracts of L cells were prepared with NE-PER Nuclear and Cytoplasmic Extraction Reagents (Pierce). Data are shown as means and s.d.

**IP/western blot analyses and binding assays.** Conditioned media for IP/western blot analyses containing full-length or various deletion mutants of IGFBPs, LRP6, Frz8CRD and Dkk1 were produced with 293 cells. Binding reactions were performed overnight at 4 °C. Immunoprecipitation was performed with Protein G-Sepharose 4 Fast Flow (Amersham). <sup>125</sup>I-labelling of IGFBP-4 and Wnt3A was performed with IODO-BEADS Iodination Reagent (Pierce). A liquid-phase binding assay was performed essentially as described<sup>19</sup>. In brief, conditioned media containing LRP6N-Myc or Frz8CRD-Myc were mixed with various concentrations of <sup>125</sup>I-labelled IGFBP-4 and incubated overnight at 4 °C. LRP6N-Myc or Frz8CRD-Myc was immunoprecipitated and the radioactivity of bound IGFBP-4 was measured after extensive washing of the Protein G-Sepharose

beads. For a competitive binding assay, conditioned media containing LRP6N-Myc or Frz8CRD-Myc were mixed with <sup>125</sup>I-labelled Wnt3A and unlabelled IGFBP-4, and incubated overnight at 4 °C. LRP6N-Myc or Frz8CRD-Myc was then immunoprecipitated and the radioactivity of bound Wnt3A was measured.

**Xenopus experiments and mouse *in situ* hybridization analysis.** Axis duplication assays, animal cap assays and *in situ* hybridization analyses in *Xenopus* were performed essentially as described<sup>20</sup>. Two independent cDNAs for XIGFBP-4, presumably resulting from pseudotetraploid genomes, were identified by 5' rapid amplification of cDNA ends (Supplementary Fig. 4a). Two different MOs targeting both of these two IGFBP-4 transcripts were designed (Gene Tools) (Supplementary Fig. 4a and Supplementary Table 2). MO-sensitive XIGFBP-4 cDNA including a 41-base-pair 5'-untranslated region (UTR) was generated by PCR. MO-resistant XIGFBP-4 cDNA (wild-type and H74P mutant) was generated by introducing five silent mutations in the MO1 target sequence and excluding the 5'-UTR (Supplementary Fig. 4a). To determine the specificity of MOs, MO-sensitive or MO-resistant XIGFBP-4-myc mRNA was injected into *Xenopus* embryos with or without MOs, and protein/mRNA expression was analysed. PCR primers and PCR conditions are listed in Supplementary Table 1. MOs and plasmid DNAs were injected at the eight-cell stage into the dorsal region of two dorsal-vegetal blastomeres fated to be heart and liver anlage. Electroporation of mRNA was performed essentially as described<sup>21</sup>. Injection of mRNA (5 ng in 5 nl of solution) into the vicinity of heart anlage and application of electric pulses were performed at stage 28. Whole-mount *in situ* hybridization analysis of murine IGFBP-4 was performed as described<sup>30</sup>.

22. He, X. *et al.* A member of the Frizzled protein family mediating axis induction by Wnt-5A. *Science* **275**, 1652–1654 (1997).
23. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
24. Mao, B. *et al.* LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. *Nature* **411**, 321–325 (2001).
25. Christian, J. L. & Moon, R. T. Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of *Xenopus*. *Genes Dev.* **7**, 13–28 (1993).
26. Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of  $\beta$ -catenin is regulated in *Xenopus* embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).
27. Kolosov, E. *et al.* Identification and characterization of embryonic stem cell-derived pacemaker and atrial cardiomyocytes. *FASEB J.* **19**, 577–579 (2005).
28. Korchynskiy, O. & ten Dijke, P. Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. *J. Biol. Chem.* **277**, 4883–4891 (2002).
29. Kishida, M. *et al.* Synergistic activation of the Wnt signaling pathway by Dvl and casein kinase I $\epsilon$ . *J. Biol. Chem.* **276**, 33147–33155 (2001).
30. Hosoda, T. *et al.* A novel myocyte-specific gene Midori promotes the differentiation of P19CL6 cells into cardiomyocytes. *J. Biol. Chem.* **276**, 35978–35989 (2001).

## LETTERS

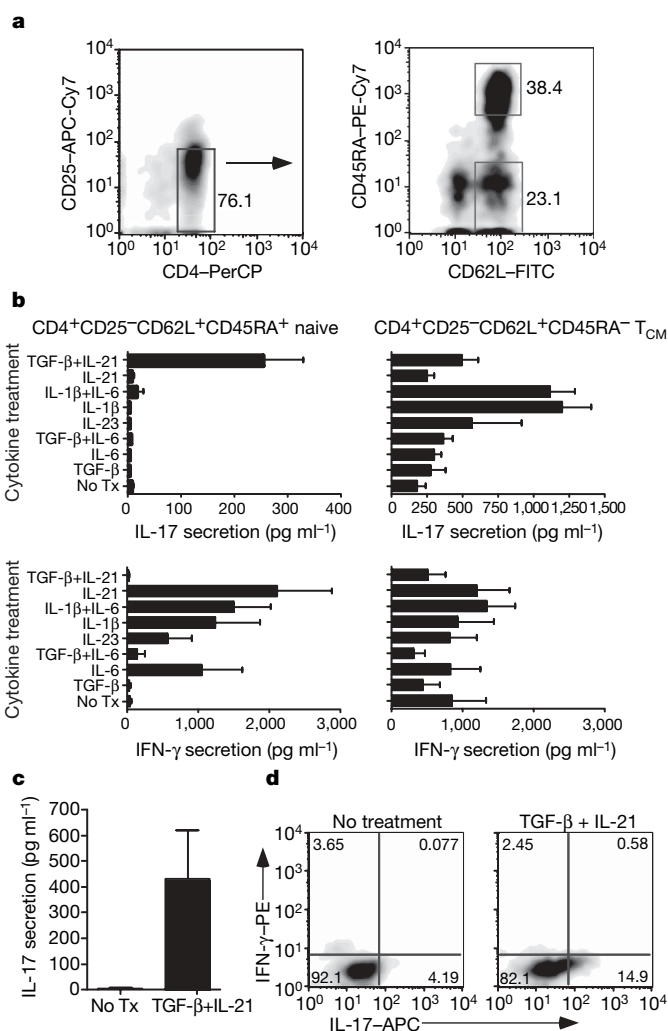
# IL-21 and TGF- $\beta$ are required for differentiation of human T<sub>H</sub>17 cells

Li Yang<sup>1\*</sup>, David E. Anderson<sup>1\*</sup>, Clare Baecher-Allan<sup>1</sup>, William D. Hastings<sup>1</sup>, Estelle Bettelli<sup>1</sup>, Mohamed Oukka<sup>1</sup>, Vijay K. Kuchroo<sup>1\*</sup> & David A. Hafler<sup>1</sup>

The recent discovery of CD4<sup>+</sup> T cells characterized by secretion of interleukin (IL)-17 (T<sub>H</sub>17 cells) and the naturally occurring regulatory FOXP3<sup>+</sup> CD4 T cell (nT<sub>reg</sub>) has had a major impact on our understanding of immune processes not readily explained by the T<sub>H</sub>1/T<sub>H</sub>2 paradigm. T<sub>H</sub>17 and nT<sub>reg</sub> cells have been implicated in the pathogenesis of human autoimmune diseases, including multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease and psoriasis<sup>1,2</sup>. Our recent data and the work of others demonstrated that transforming growth factor- $\beta$  (TGF- $\beta$ ) and IL-6 are responsible for the differentiation of naive mouse T cells into T<sub>H</sub>17 cells, and it has been proposed that IL-23 may have a critical role in stabilization of the T<sub>H</sub>17 phenotype<sup>3–5</sup>. A second pathway has been discovered in which a combination of TGF- $\beta$  and IL-21 is capable of inducing differentiation of mouse T<sub>H</sub>17 cells in the absence of IL-6 (refs 6–8). However, TGF- $\beta$  and IL-6 are not capable of differentiating human T<sub>H</sub>17 cells<sup>2,9</sup> and it has been suggested that TGF- $\beta$  may in fact suppress the generation of human T<sub>H</sub>17 cells<sup>10</sup>. Instead, it has been recently shown that the cytokines IL-1 $\beta$ , IL-6 and IL-23 are capable of driving IL-17 secretion in short-term CD4<sup>+</sup> T cell lines isolated from human peripheral blood<sup>11</sup>, although the factors required for differentiation of naive human CD4 to T<sub>H</sub>17 cells are still unknown. Here we confirm that whereas IL-1 $\beta$  and IL-6 induce IL-17A secretion from human central memory CD4<sup>+</sup> T cells, TGF- $\beta$  and IL-21 uniquely promote the differentiation of human naive CD4<sup>+</sup> T cells into T<sub>H</sub>17 cells accompanied by expression of the transcription factor RORC2. These data will allow the investigation of this new population of T<sub>H</sub>17 cells in human inflammatory disease.

To better understand regulation of IL-17A secretion from human CD4<sup>+</sup> T cells, we used a strategy that would allow us to evaluate the effects of various combinations of cytokine on expansion of T<sub>H</sub>17 cells from memory T cells versus differentiation of naive CD4<sup>+</sup> lymphocytes into T<sub>H</sub>17 cells. Specifically, we used high-speed flow cytometry for sorting these two distinct populations of CD4<sup>+</sup> T cells from the peripheral blood of healthy subjects: CD4<sup>+</sup>CD25<sup>–</sup>CD62L<sup>+</sup>CD45RA<sup>hi</sup> cells highly enriched for naive T cells and CD4<sup>+</sup>CD25<sup>+</sup>CD62L<sup>+</sup>CD45RA<sup>–</sup> cells enriched for central memory T cells (T<sub>CM</sub>; Fig. 1a). All cells enriched for a naive or a central memory phenotype expressed the chemokine receptor CCR7 (data not shown). These two T cell populations were then stimulated with plate-bound anti-CD3 and soluble anti-CD28 monoclonal antibodies for 7 days in serum-free medium containing different combinations of cytokines implicated in CD4<sup>+</sup> T<sub>H</sub>17 cell differentiation.

As reported previously, the cytokine IL-1 $\beta$  induced the greatest amount of IL-17A secretion from T<sub>CM</sub> (Fig. 1b). The addition of IL-6 alone had little effect on induction of IL-17A, and when added with IL-1 $\beta$  had no additive or synergistic effect on IL-17A production.



**Figure 1 | TGF- $\beta$  and IL-21 promote T<sub>H</sub>17 differentiation from naive CD4<sup>+</sup> T cells.** **a**, CD4<sup>+</sup> T cells were sorted into populations enriched for naive or central memory T helper cells. **b**, IFN- $\gamma$  and IL-17A secretion is shown from T cells stimulated for 7 days in the presence of the indicated cytokines. Standard deviation using T cells from three unrelated subjects is represented. TGF- $\beta$ - and IL-21-induced IL-17 secretion is highly significant ( $P < 0.01$ ). Tx, treatment. **c**, IL-17 secretion from naive T cells from six different donors is represented (mean  $\pm$  s.e.m.). **d**, Intracellular expression of IL-17 and IFN- $\gamma$  from one of five experiments is shown.

<sup>1</sup>Division of Molecular Immunology, Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.

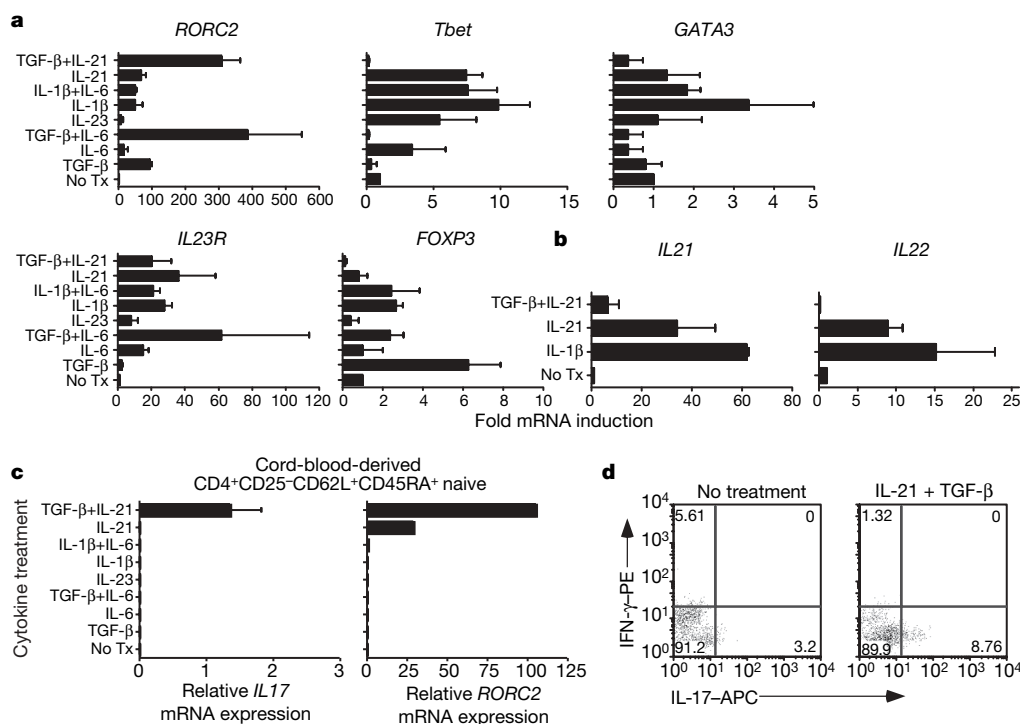
Addition of IL-23 was also able to modestly enhance IL-17A secretion from  $T_{CM}$ . However, IL-1 $\beta$  alone or together with IL-6 failed to induce IL-17A secretion from naive  $CD4^+$  T cells. In marked contrast, a combination of TGF- $\beta$  and IL-21 was uniquely able to induce  $T_{H17}$  differentiation. Whereas IL-21, IL-1 $\beta$  or IL-6 induced significant amounts of interferon- $\gamma$  (IFN- $\gamma$ ) secretion from naive T cells, the addition of TGF- $\beta$  with IL-21 suppressed IFN- $\gamma$  secretion and induced differentiation of  $T_{H17}$  cells. Whereas we did observe variability in the extent of  $T_{H17}$  differentiation among unrelated healthy donors (Fig. 1c), we always observed induction of IL-17A after differentiation in the presence of TGF- $\beta$  with IL-21. Intracytoplasmic staining demonstrated, in agreement with enzyme-linked immunosorbent assay (ELISA) results, that the combination of TGF- $\beta$  and IL-21 differentiated  $CD4^+$  T cells that secreted only IL-17A and no IFN- $\gamma$  (Fig. 1d). A recent study<sup>12</sup> has demonstrated that the pathogenicity of mouse IL-17-secreting T cells is influenced by whether they also secrete IL-10. Using intracytoplasmic staining, we failed to observe any IL-10 when naive  $CD4^+$  T cells were differentiated in the presence of TGF- $\beta$  and IL-21 (data not shown).

RORC2 is the human homologue of mouse ROR $\gamma$ t—a transcription factor critical for the differentiation of mouse IL-17-secreting T cells. Thus, to understand the molecular mechanisms involved in differentiation of human  $T_{H17}$  cells, we used quantitative PCR with reverse transcription (RT-PCR) to evaluate messenger RNA levels of RORC2 and other molecules implicated in  $T_{H17}$  differentiation. The combination of TGF- $\beta$  and IL-21 induced high levels of RORC2 (Fig. 2a), consistent with their ability to induce IL-17A secretion from naive human  $CD4^+$  T cells. It was of particular interest that the combination of TGF- $\beta$  and IL-6 that induces  $T_{H17}$  differentiation in mouse T cells also induced expression of RORC2 in naive human  $CD4$  cells. Because this combination of cytokines did not, however, induce IL-17A secretion, these data indicate that expression of RORC2 in humans is not in itself sufficient to induce IL-17 production, and that another as yet unidentified transcription factor in combination with RORC2, perhaps the human homologue of mouse

ROR $\alpha$  (ref. 13), may be required to induce IL-17A-secreting  $T_{H17}$  cells.

T-bet (also known as TBX21) is the master regulator for  $T_{H1}$  cells secreting IFN- $\gamma$ , whereas GATA3 induces  $T_{H2}$ , IL-4 secreting  $CD4$  cells. Messenger RNA expression levels of *Tbet* were highly concordant with amounts of IFN- $\gamma$  secretion and were consistent with our findings that whereas TGF- $\beta$  and IL-21 induce  $T_{H17}$  cell differentiation with RORC2 expression, TGF- $\beta$  suppresses the induction of T-bet by IL-21. Similarly, there was no induction of GATA3 with TGF- $\beta$  and IL-21. The cytokines IL-6, IL-21 and IL-1 $\beta$  but not TGF- $\beta$  induced IL-23 receptor upregulation in stimulated naive  $CD4^+$  T cells. We also examined the expression of the  $T_{reg}$  transcription factor FOXP3. As has been previously reported in both mouse and human systems, FOXP3 was induced by TGF- $\beta$ . Induction of FOXP3 was inhibited by IL-6 and to a greater extent by IL-21—transcription factors that induce RORC2. Thus, although the induction of RORC2 and FOXP3 transcription factors was highly similar between mouse and human naive  $CD4$  cells, the induction of IL-17A by IL-6 in combination with TGF- $\beta$  is discordant between the species.

We and others have shown previously that IL-21 secreted by mouse  $CD4^+$  T cells can induce the secretion of IL-21 in an autocrine amplification loop<sup>6–8,14</sup>. Thus, we examined whether human IL-21 similarly induced IL-21 secretion from naive  $CD4^+$  T cells; we also evaluated the effects of a combination of TGF- $\beta$  with IL-21 and IL-1 $\beta$ , given the ability of these cytokines to induce IL-17 from naive and central memory  $CD4^+$  T cells. Consistent with results in mice, IL-21 upregulated IL-21, although IL-1 $\beta$  induced even greater amounts of *IL21* mRNA (Fig. 2b). In contrast to what has been observed in mice, IL-21 also increased *IL22* mRNA levels in naive  $CD4^+$  T cells in the absence of any exogenous IL-23. However, TGF- $\beta$  suppressed the expression of *IL21* and *IL22* mRNA induced by IL-21 (Fig. 2b). These data further highlight similar yet subtle differences between human and mouse  $CD4^+$  T cells, because whereas IL-21 induces *IL21* and *IL22*, differentiation to  $T_{H17}$  cells with TGF- $\beta$  inhibits the expression of these cytokines.



**Figure 2 | TGF- $\beta$  and IL-21 induce RORC2 in naive  $CD4^+$  T cells.** **a**, mRNA expression levels (fold-induction relative to T cells without exogenous cytokines) of *RORC2*, *Tbet*, *GATA3*, *IL23R* and *FOXP3* are shown after naive T cells were differentiated as indicated. **b**, Fold-induction  $\pm$  s.e.m. ( $n = 3$ ) of *IL21* and *IL22* are represented. **c**, Mean expression  $\pm$  s.e.m. of *IL17A* and

*RORC2* are shown ( $n = 3$ ) for naive T cells obtained from cord blood. **d**, Shown is intracytoplasmic staining of IL-17 and IFN- $\gamma$  from cord blood naive T cells after 7 days of differentiation. Similar results were seen in another independent assay.



To confirm the unique function of TGF- $\beta$  and IL-21 in the differentiation of T<sub>H</sub>17 cells from naive human CD4<sup>+</sup> T cells, we sorted CD4<sup>+</sup>CD25<sup>-</sup>CD62L<sup>+</sup>CD45RA<sup>hi</sup> cells from human cord blood. As expected, a higher proportion of CD4 cells in the cord blood exhibited this naive phenotype relative to peripheral blood obtained from healthy adult subjects (data not shown). TGF- $\beta$  and IL-21 induced the upregulation of *IL17A* and *RORC2* mRNA (Fig. 2c). Although IL-21 alone modestly induced *RORC2*, only TGF- $\beta$  and IL-21 were able to induce *IL17A* mRNA. When given a very strong *in vitro* stimulus, naive CD4<sup>+</sup> T cells sorted from cord blood secreted IL-17A protein (Fig. 2d). These data further indicate that TGF- $\beta$  and IL-21 are critical in the differentiation of both human and mouse T<sub>H</sub>17 cells.

Collectively, our data refine and extend our understanding of the regulation of IL-17A secretion from human CD4<sup>+</sup> T cells and define the conditions required for human T<sub>H</sub>17 cell differentiation. We confirm recent reports that IL-1 $\beta$  together with IL-6 (ref. 9) or IL-23 (ref. 2) can induce IL-17A secretion, which is most apparent in the human memory CD4<sup>+</sup> T cell subset. A combination of TGF- $\beta$  plus IL-21 is required for the differentiation of T<sub>H</sub>17 cells from naive T cells. It is important to note that to observe T<sub>H</sub>17 differentiation from naive human CD4<sup>+</sup> T cells, appropriate amounts of both IL-21 and TGF- $\beta$  are needed: addition of IL-21 alone must be sufficient to upregulate IFN- $\gamma$  secretion, and the amount of TGF- $\beta$  added must inhibit IL-21-induced IFN- $\gamma$  secretion (see Methods for details). T<sub>H</sub>17 cells differentiated under these conditions are also notable for secretion of IL-17A in the absence of IFN- $\gamma$ . In summary, our data together with previous reports in humans suggest that IL-1 $\beta$  and IL-6 induced during the early stages of an inflammatory response may act on memory T cells to promote IL-17 and IL-21 secretion, with induced IL-21 able to synergize with TGF- $\beta$  to promote differentiation of T<sub>H</sub>17 cells from naive CD4<sup>+</sup> T cells. IL-23 may serve to expand or stabilize the phenotype of previously differentiated T<sub>H</sub>17 cells. These experiments allow for the characterization of human inflammatory T<sub>H</sub>17 responses associated with infection and autoimmune diseases.

## METHODS SUMMARY

**Cell sorting.** PBMCs were obtained from the peripheral blood of healthy subjects or from cord blood (AllCells) in compliance with institutional review board (IRB) protocols. CD4<sup>+</sup> T cells were subsequently isolated by negative selection using magnetic beads (Miltenyi Biotec). Naive (CD25<sup>-</sup>CD62L<sup>+</sup>CD45RA<sup>hi</sup>) and central memory (CD25<sup>+</sup>CD62L<sup>+</sup>CD45RA<sup>+</sup>) CD4<sup>+</sup> T cells were obtained by staining with the following antibodies: CD62L-FITC, CD4-PerCP, CD45RA-PE-Cy7, CD25-APC-Cy7 (BD Pharmingen) and were sorted on a FACS Aria (BD Biosciences).

**Differentiation assays.** Naive or central memory CD4<sup>+</sup> T cells were stimulated with plate-bound anti-CD3 and soluble CD28 monoclonal antibodies in serum-free X-VIVO15 medium (Biowhittaker) and cytokines (IL-6, 25 ng ml<sup>-1</sup>; TGF- $\beta$ 1, 5 ng ml<sup>-1</sup>; IL-1 $\beta$ , 12.5 ng ml<sup>-1</sup>; IL-21, 25 ng ml<sup>-1</sup>; IL-23, 25 ng ml<sup>-1</sup>) for a period of 7 days, at which point supernatants were collected and tested by ELISA for IFN- $\gamma$  (BD Biosciences) or IL-17A (eBioscience) using paired

antibodies. We have observed that concentrations of TGF- $\beta$  ranging from 0.1 ng ml<sup>-1</sup> to 10 ng ml<sup>-1</sup> in the presence of IL-21 promote T<sub>H</sub>17 differentiation, whereas 50 ng ml<sup>-1</sup> TGF- $\beta$  suppresses differentiation. Intracytoplasmic staining was performed using standard methodologies and anti-IL-17-APC (R&D Systems) and anti-IFN- $\gamma$ -PE or anti-IL-10-PE (BD Biosciences) antibodies.

**Real-time PCR.** All primers and probes were obtained from Applied Biosystems and used according to standard methodologies.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 17 October 2007; accepted 24 April 2008.

Published online 11 May 2008.

1. Kikly, K., Liu, L., Na, S. & Sedgwick, J. D. The IL-23/Th(17) axis: therapeutic targets for autoimmune inflammation. *Curr. Opin. Immunol.* **18**, 670–675 (2006).
2. Wilson, N. J. *et al.* Development, cytokine profile and function of human interleukin 17-producing helper T cells. *Nature Immunol.* **8**, 950–957 (2007).
3. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T<sub>H</sub>17 and regulatory T cells. *Nature* **441**, 235–238 (2006).
4. Mangan, P. R. *et al.* Transforming growth factor- $\beta$  induces development of the T<sub>H</sub>17 lineage. *Nature* **441**, 231–234 (2006).
5. Veldhoen, M., Hocking, R. J., Atkins, C. J., Locksley, R. M. & Stockinger, B. TGF $\beta$  in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. *Immunity* **24**, 179–189 (2006).
6. Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T<sub>H</sub>17 cells. *Nature* **448**, 484–487 (2007).
7. Nurieva, R. *et al.* Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature* **448**, 480–483 (2007).
8. Zhou, L. *et al.* IL-6 programs T<sub>H</sub>17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nature Immunol.* **8**, 967–974 (2007).
9. Acosta-Rodriguez, E. V., Napolitani, G., Lanzavecchia, A. & Sallusto, F. Interleukins 1 $\beta$  and 6 but not transforming growth factor- $\beta$  are essential for the differentiation of interleukin 17-producing human T helper cells. *Nature Immunol.* **8**, 942–949 (2007).
10. Evans, H. G., Suddason, T., Jackson, I., Taams, L. S. & Lord, G. M. Optimal induction of T helper 17 cells in humans requires T cell receptor ligation in the context of Toll-like receptor-activated monocytes. *Proc. Natl Acad. Sci. USA* **104**, 17034–17039 (2007).
11. Laurence, A. & O'Shea, J. J. T<sub>H</sub>17 differentiation: of mice and men. *Nature Immunol.* **8**, 903–905 (2007).
12. McGeachy, M. J. *et al.* TGF- $\beta$  and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain T<sub>H</sub>17 cell-mediated pathology. *Nature Immunol.* **8**, 1390–1397 (2007).
13. Yang, X. O. *et al.* T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR $\alpha$  and ROR $\gamma$ . *Immunity* **28**, 29–39 (2008).
14. Wei, L., Laurence, A., Elias, K. M. & O'Shea, J. J. IL-21 is produced by Th17 cells and drives IL-17 production in a STAT3-dependent manner. *J. Biol. Chem.* **282**, 34605–34610 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Kozoriz for assistance with flow cytometric cell sorting. Supported by grants from the NIH (D.A.H., V.K.K.), the National Multiple Sclerosis Society (D.A.H., V.K.K.), the Juvenile Diabetes Research Foundation (D.A.H., V.K.K.), and the American Cancer Society (D.E.A.). D.A.H. and V.K.K. are recipients of a Javitz Investigator award from the NIH.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.E.A. ([danderson@rics.bwh.harvard.edu](mailto:danderson@rics.bwh.harvard.edu)).

## METHODS

**Differentiation assays.** Plates (96-well U-bottom) were coated with  $1.5 \mu\text{g ml}^{-1}$  anti-CD3 monoclonal antibody (eBioscience, clone OKT3) in a volume of 50  $\mu\text{l}$  of PBS and were incubated overnight (16 h) at  $4^\circ\text{C}$ . For T cells isolated from cord blood, a 96-well plate was pre-coated overnight at  $4^\circ\text{C}$  with  $3 \mu\text{g ml}^{-1}$  anti-CD3 monoclonal antibody (UCHT1, BD Biosciences). Antibody solution was then removed, plates were rinsed once with serum-free X-VIVO 15 medium (Biowhittaker), and naive or central memory  $\text{CD4}^+$  T cells ( $5 \times 10^4$  per well) were stimulated in serum-free X-VIVO15 medium with soluble CD28 (BD Biosciences, clone 28.2) monoclonal antibody ( $1.0 \mu\text{g ml}^{-1}$ ) and cytokines (IL-6,  $25 \text{ ng ml}^{-1}$ ; TGF- $\beta$ 1,  $5 \text{ ng ml}^{-1}$ ; IL-1 $\beta$ ,  $12.5 \text{ ng ml}^{-1}$ ; IL-21,  $25 \text{ ng ml}^{-1}$ ; IL-23,  $25 \text{ ng ml}^{-1}$ ) for a period of 7 days in the absence of IL-2. The cytokines IL-6, IL-1 $\beta$ , IL-23 and TGF $\beta$ 1 (catalogue number 240-B-002) were all obtained from R&D systems. The TGF $\beta$ 1 was not acid-treated before addition. IL-21 was purchased from Cell Sciences (catalogue number CRI 172A). Supernatants were collected and tested by ELISA for IFN- $\gamma$  (Endogen) or IL-17A (human IL-17A ELISA kit from eBioscience or human IL-17 duoset from R&D systems) using paired antibodies. We have observed that concentrations of TGF- $\beta$ 1 ranging from  $0.1 \text{ ng ml}^{-1}$  to  $10 \text{ ng ml}^{-1}$  in the presence of IL-21 promote  $\text{T}_\text{H}17$  differentiation, with lower doses of TGF $\beta$ 1 associated with induction of a higher proportion of IL-17-producing cells than higher doses; use of  $50 \text{ ng ml}^{-1}$  TGF- $\beta$ 1 suppresses differentiation and IL-17 secretion. Intracytoplasmic staining was performed using standard methodologies and anti-IL-17-APC (R&D Systems) and anti-IFN- $\gamma$ -PE or anti-IL-10-PE (BD Biosciences) antibodies. Cells were stimulated with phorbol 12-myristate 13-acetate (PMA;  $10 \text{ ng ml}^{-1}$ ), ionomycin ( $0.5 \mu\text{g ml}^{-1}$ ) and Golgistop for 5 h at  $37^\circ\text{C}$  before intracellular staining.

**Real-time PCR.** After removing supernatants from wells in differentiation assays, 250  $\mu\text{l}$  per well of lysis buffer was added, at which point RNA was isolated using a RNeasy mini kit (Qiagen). Total RNA was converted to complementary DNA using Taqman Reverse transcription reagents (Applied Biosystems). Quantitative PCR was performed using a 7500 Fast Real-time PCR system (Applied Biosystems). All primers and probes were obtained from Applied Biosystems and used according to standard methodologies.

# Positive feedback sharpens the anaphase switch

Liam J. Holt<sup>1</sup>, Andrew N. Krutchinsky<sup>2</sup> & David O. Morgan<sup>1</sup>

At the onset of anaphase, sister-chromatid cohesion is dissolved abruptly and irreversibly, ensuring that all chromosome pairs disjoin almost simultaneously. The regulatory mechanisms that generate this switch-like behaviour are unclear. Anaphase is initiated when a ubiquitin ligase, the anaphase-promoting complex (APC), triggers the destruction of securin, thereby allowing separase, a protease, to disrupt sister-chromatid cohesion<sup>1–4</sup>. Here we demonstrate that the cyclin-dependent kinase 1 (Cdk1)-dependent phosphorylation of securin near its destruction-box motif inhibits securin ubiquitination by the APC. The phosphatase Cdc14 reverses securin phosphorylation, thereby increasing the rate of securin ubiquitination. Because separase is known to activate Cdc14 (refs 5 and 6), our results support the existence of a positive feedback loop that increases the abruptness of anaphase. Consistent with this model, we show that mutations that disrupt securin phosphoregulation decrease the synchrony of chromosome segregation. Our results also suggest that coupling securin degradation with changes in Cdk1 and Cdc14 activities helps coordinate the initiation of sister-chromatid separation with changes in spindle dynamics.

Securin is known to integrate multiple signalling pathways to delay anaphase in response to perturbations such as DNA damage<sup>7,8</sup>. We hypothesized that securin might also receive regulatory inputs that make separase activation more switch-like. To address this possibility, we used mass spectrometry to analyse the phosphorylation state of securin from mitotic budding yeast cells (Fig. 1a)<sup>9</sup>. We identified six phosphorylation sites, including four of the five sites corresponding to the consensus sequence of Cdk1 (S/T\*-P) (see also Supplementary Fig. 1). Three Cdk1 sites were found in securin's carboxy (C)-terminal domain, which is known to interact with separase<sup>10</sup>; mutations at these sites are known to modulate securin's ability to bind separase and import it into the nucleus<sup>11</sup>. The fourth Cdk1 phosphopeptide, near the amino (N) terminus of securin, has not been characterized.

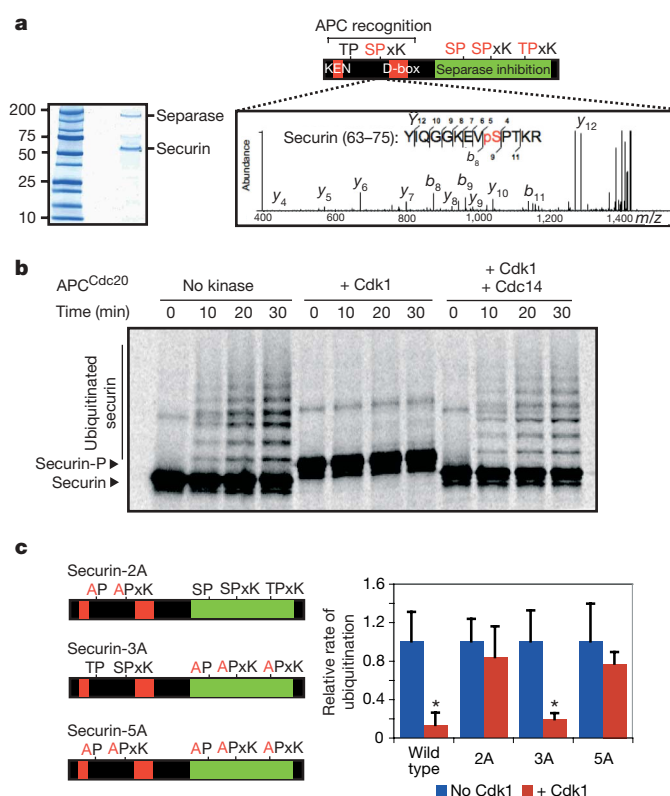
The N-terminal Cdk1 site is near the destruction-box of securin, a motif important for recognition of substrates by the APC<sup>2–4,12</sup>. To assess whether phosphorylation of this site affects securin ubiquitination by the APC, we tested the ability of purified APC to ubiquitinate purified securin before and after phosphorylation *in vitro* by Cdk1. We found that if securin was phosphorylated by Cdk1, the rate of ubiquitination by APC<sup>Cdc20</sup> (the APC-activator complex that controls anaphase onset) was reduced tenfold (Fig. 1b, c) and the rate of ubiquitination by APC<sup>Cdh1</sup> was reduced fivefold (Supplementary Fig. 2).

This regulatory mechanism may be conserved: *Drosophila* and human homologues of securin have S/T-P motifs near their destruction-boxes, and both were ubiquitinated less efficiently by the APC after phosphorylation by Cdk1 (Supplementary Fig. 3).

Cdc14 is a phosphatase that removes phosphates from Cdk1 substrates during anaphase and late mitosis<sup>4,13–16</sup>. We found that Cdc14 removed Cdk1-dependent phosphates from securin, as judged by the loss of a gel-mobility shift that results from phosphorylation (Fig. 1b;

Supplementary Fig. 2a). Dephosphorylation reactivated securin as an efficient APC substrate.

To confirm that regulation depends on the canonical Cdk1 sites, we mutagenized securin to remove the N-terminal Cdk1 sites



**Figure 1 | Cdk1 and Cdc14 control the phosphorylation state of securin near its destruction-box and modulate the rate of securin ubiquitination.**

**a**, Securin was purified from a benomyl-arrested strain carrying *SECURIN-3FLAG-6HIS* (left panel), and tandem mass spectrometry analysis resulted in the detection of ten phosphopeptides (Supplementary Fig. 1), including four peptides phosphorylated at Cdk1 consensus sites (red). One of these sites lies next to the securin destruction-box motif (the MS<sup>3</sup> spectrum of the phosphopeptide is shown to the right, with the phosphopeptide sequence most consistent with the data).

**b**, [<sup>35</sup>S]methionine-labelled securin was produced by translation *in vitro* and incubated with purified APC<sup>Cdc20</sup> and other ubiquitination components. Before the reactions, as indicated, some samples were incubated with either purified Clb2–Cdk1 or with Clb2–Cdk1 and then Cdc14 sequentially. The substrate was purified away from kinase and phosphatase before addition of APC<sup>Cdc20</sup>.

**c**, Quantification of experiments conducted as in **b** using either wild-type securin, securin-2A (T27A, S71A), securin-3A (S277A, S292A, T304A) or securin-5A as substrate. Error bars, s.d. from three experiments. Asterisks, *P* < 0.05 (*t*-test) when comparing results with and without kinase.

<sup>1</sup>Departments of Physiology, Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. <sup>2</sup>Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158, USA.



(securin-2A), the three C-terminal sites (securin-3A) or all five sites (securin-5A). Securin-2A and securin-5A were still recognized by the APC after incubation with Cdk1 (Fig. 1c and Supplementary Fig. 2b), whereas securin-3A ubiquitination was inhibited by Cdk1 (Fig. 1c). We conclude that the ability of Cdk1 and Cdc14 to control securin ubiquitination depends on one or both of the N-terminal Cdk1 sites.

Current models of anaphase regulation do not provide many opportunities for switch-like behaviour; in particular there is no known positive feedback in this system (Fig. 2a)<sup>17,18</sup>. It is therefore interesting that separase, in addition to cleaving cohesin, initiates the activation of Cdc14 in early anaphase<sup>5,6</sup>. Thus, our observation that Cdc14 can dephosphorylate securin, and thereby increase its efficiency as an APC substrate, leads to the potential for a positive feedback loop that is predicted to make anaphase more switch-like (Fig. 2b).

APC<sup>Cdc20</sup> activation triggers the destruction of cyclins as well as securin (Fig. 2a, b). Thus, the kinases responsible for securin phosphorylation are partly inactivated at anaphase initiation, providing an additional mechanism by which securin destruction is promoted, and potentially increasing the robustness of this switch.

We explored a simple mathematical version of this model (Fig. 2c; see Supplementary Fig. 4 for details) by numerically determining the

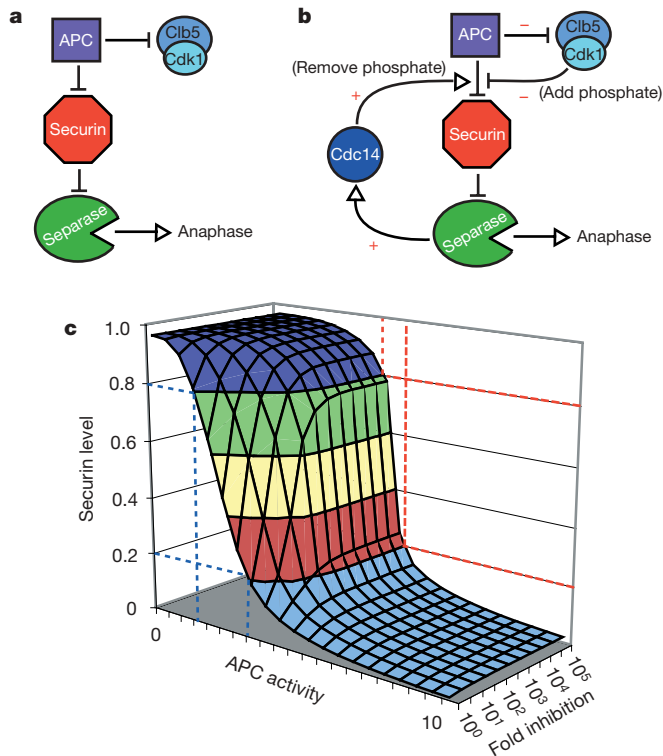
steady-state levels of securin as a function of APC activity under a range of parameters. We did not include the effects of the APC on Cdk activity in the model, as it did not significantly affect the outcome of the simulations (Supplementary Fig. 5). To explore how the modulation of securin destruction might affect the abruptness of anaphase, we varied the effect of securin phosphorylation on the rate of its destruction. As we increased the inhibitory effect of phosphorylation from 1- to 100-fold, securin degradation became more switch-like (Fig. 2c).

To investigate the relevance of our model *in vivo*, we created a yeast strain in which we specifically disrupted the putative positive feedback loop by replacing the securin gene with a mutant gene encoding securin-2A. This strain was viable at both 30 °C (the restrictive temperature for a *securinΔ* strain) and 37 °C, suggesting that securin-2A is functional in its ability to bind and fold separase. We measured the timing of securin-2A degradation in cells released from a G1 arrest and found no significant difference between the mutant and wild-type securin (Supplementary Fig. 6). We suspected, however, that asynchrony in the cell population in these experiments prevented the detection of small differences in the timing and abruptness of securin destruction. We therefore attempted to develop microscopic methods to measure securin destruction in single cells. Green fluorescent protein (GFP)-tagged securin was undetectable when expressed at normal levels, and overexpressed securin-GFP was detectable but delayed cells in late mitosis. It was possible, however, to assess the relative stabilities of securin and securin-2A in late mitosis by simultaneously overexpressing and analysing the levels of both proteins, each tagged with a different GFP variant. We found that mitotic securin-2A levels were lower than those of wild-type securin (Supplementary Fig. 7).

To analyse the consequences of securin phosphoregulation in more detail, we developed methods for the analysis of spindle dynamics and the abruptness of anaphase onset in single cells. To quantify spindle length, we tagged the spindle-pole body protein, Spc42, with dsRed. To visualize the rate and abruptness of chromosome segregation, we integrated *LAC* and *TET* operator arrays at two independent loci, *TRP1* and *URA3*, respectively, and expressed GFP-tagged Lac and Tet repressors in the cell to generate fluorescent marks at these loci<sup>19,20</sup>. Chromosomes IV (*TRP1-LACO*) and V (*URA3-TETO*) were thereby marked at loci that are close to the centromere (12 and 35 kilobases, respectively) but far enough away that they are not resolved until cohesin is cleaved by separase. One GFP-labelled locus was consistently brighter than the other, and we found that this dot disappeared upon addition of tetracycline, suggesting that the bright dot labels chromosome V (Supplementary Fig. 8).

We used a spinning-disk confocal microscope to obtain movies of anaphase at a time resolution of 10 s. These movies allowed us to determine the time lag between the segregation of one chromosome and that of a second, providing a direct measure of the abruptness of anaphase. In addition, we measured the rates and coordination of spindle elongation and chromosome segregation.

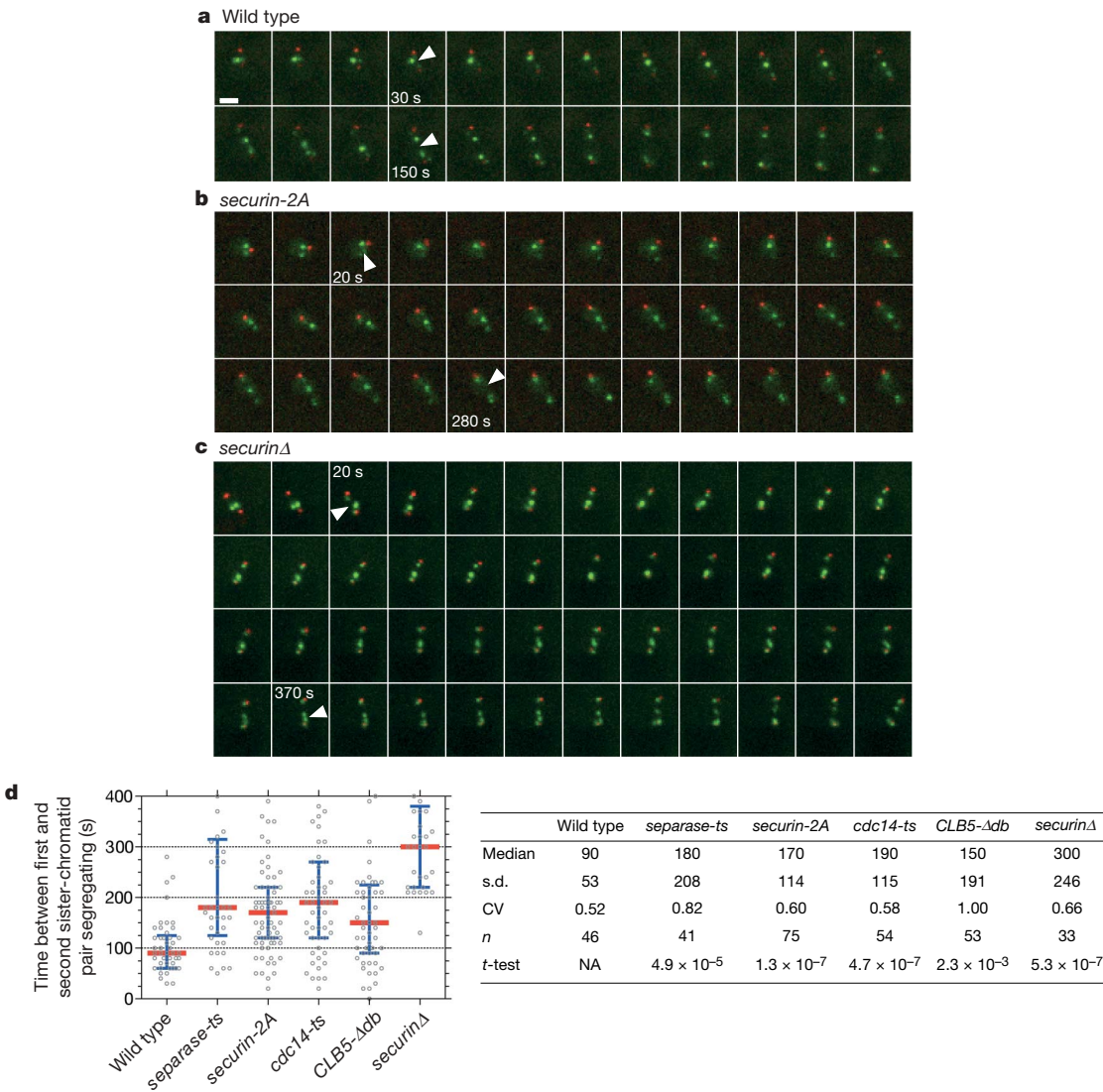
If there is any variation in the inherent sensitivity of each chromosome to separase activity—that is, if different chromatid pairs disjoin at different thresholds of separase activity—then the time lag between segregation of one pair of sister chromatids and another is a measure of the abruptness of separase activation (Supplementary Fig. 9). Indeed, we found that there was a reproducible time lag between the segregation of the two labelled chromosomes in our strain. In addition, the order of chromosome segregation was invariant: in 100% of cells, the bright dot on chromosome V segregated after the dot on chromosome IV. Quantification of 40 wild-type anaphases (Fig. 3a, d) revealed that the median time between the two chromosome segregation events ( $t_{\text{seg}}$ ) was 90 s, with most cells clustered tightly around this value. This abrupt timing of anaphase depends on rapid activation of separase: a mutant with reduced separase function (*esp1-1*) displayed a less synchronous anaphase (median  $t_{\text{seg}}$  = 180 s).



**Figure 2 | Modulation of securin ubiquitination by Cdk1 and Cdc14 gives rise to a potential positive feedback loop in the anaphase regulatory network.** **a**, Simplified previous model for anaphase regulation in *S. cerevisiae*: the APC targets securin and some mitotic cyclins for destruction, which liberates separase to cleave cohesin and initiate anaphase. **b**, Modified model of anaphase control: Cdk1 phosphorylates securin, reducing the rate at which it is ubiquitinated by the APC. On accumulation of high levels of APC activity, some securin is destroyed, releasing a small amount of separase. Separase activates Cdc14, which dephosphorylates securin, increasing the rate at which it is ubiquitinated by the APC. Concomitantly, some mitotic cyclins are destroyed, reducing Cdk1 activity and allowing Cdc14 to dephosphorylate securin more efficiently. **c**, Set of numerical solutions to a simplified mathematical model of the network in **b** (but excluding cyclin degradation by the APC), showing the steady-state levels of securin remaining (y-axis) as APC activity is varied (x-axis) when securin ubiquitination is inhibited to various degrees by phosphorylation (z-axis) (see Supplementary Fig. 3 for details of model).

If the positive feedback loop proposed in Fig. 2b is operating *in vivo*, then disruption of this loop should decrease the abruptness of anaphase. Indeed, the *securin-2A* mutant strain had a  $t_{\text{seg}}$  of 170 s, almost double that of wild-type cells (Fig. 3b, d). Our model (Fig. 2b) predicts that positive feedback should be weakened by reducing Cdc14 activity or increasing Cdk1 activity. Consistent with this possibility, we observed less anaphase synchrony in *cdc14-1* mutant cells ( $t_{\text{seg}}$  = 190 s) and in cells expressing Clb5- $\Delta$ db, a partly stabilized cyclin mutant that lacks its destruction-box and thus generates high levels of Cdk1 activity in anaphase ( $t_{\text{seg}}$  = 150 s) (ref. 21). Defects in Cdc14 or Clb5 affect numerous processes in addition to securin degradation. Nevertheless, when taken together with the effects of the securin-2A mutant, these results support the hypothesis that phosphorylation of the N-terminal Cdk1 sites of securin creates a positive feedback loop that makes anaphase more switch-like.

We also analysed cells carrying a deletion of securin, the loss of which reduces separase activity and partly uncouples anaphase initiation from APC activity. Securin deletion dramatically reduced the synchrony of sister separation ( $t_{\text{seg}}$  = 300 s; Fig. 3c, d), arguing that a major role for securin is to increase the abruptness of the metaphase-to-anaphase transition. A second possible function for phosphoregulation of securin is the coordination of separase activation with changes in anaphase spindle behaviour. After APC<sup>Cdc20</sup> activation, partial cyclin destruction, together with Cdc14 activation by separase, results in dephosphorylation of a subset of Cdk1 substrates that trigger important changes in the behaviour of spindle and kinetochore proteins<sup>4,22–25</sup>. These changes must be coordinated with the cleavage of cohesin. We propose that this coordination depends in part on phosphoregulation of securin, which we predict delays securin degradation relative to that of cyclins, allowing more robust dephosphorylation of Cdk1



**Figure 3 | Modulation of securin ubiquitination by Cdk1 is required for an abrupt anaphase.** **a**, A strain containing *SPC42-dsRed:NAT*; *trp1::256×LACO:TRP1*; *his3::CUP1-GFP-LACI:HIS3*; *ura3::112×TETO:URA3*; *leu2::TETR-GFP:LEU2* has red spindle poles and GFP dots marking chromosomes IV and V. Each panel in this montage represents 10 s in a time-lapse sequence during progression through anaphase. Arrowheads indicate the first and second chromosome separation. Each panel is a maximum intensity projection of a 7-μm stack. Note that the daughter spindle pole can be difficult to visualize owing to slow maturation of the dsRed fluorophore. Scale bar, 2 μm. **b**, Anaphase onset in a

*securin-2A* strain. **c**, Anaphase onset in a *securinΔ* strain. **d**, Anaphase synchrony (time between the first and second chromosomes segregating) in approximately 40 cells of wild-type, separase mutant (*esp1-1*), *securin-2A*, *cdc14-1*, *CLB5-Δdb* and *securinΔ* strains. The time of sister-chromatid separation was defined as the first time point when chromatids separated and did not re-anneal in subsequent frames. The red line is the median and the blue bars indicate the inter-quartile range. The table below shows the median, s.d. and coefficient of variation (CV; s.d. divided by the mean) for each strain. NA, not applicable. The movies corresponding to panels **a–c** are found in Supplementary Movies 1–3, respectively.

substrates by the time cohesin cleavage begins. We also hypothesize that switch-like activation of separase generates switch-like activation of Cdc14, resulting in robust and coordinated dephosphorylation of the many factors controlling anaphase spindle behaviour.

We therefore investigated anaphase dynamics in wild-type cells and in cells with perturbed securin regulation. In wild-type cells (Fig. 4a), anaphase was robust and stereotypical, as illustrated by the spindle-pole and chromosome tracking data in Fig. 4b. Each pair of sister chromatids segregated rapidly and unidirectionally (quantified in Fig. 4c), and the spindle elongated robustly in two phases: a rapid initial phase (spindle rate I) and a slower second phase (spindle rate II), as described previously<sup>19</sup>. Spindle elongation initiated at the same time as chromosome disjunction (Fig. 4a), illustrating the tight temporal coordination of cohesin cleavage with changes in spindle dynamics.

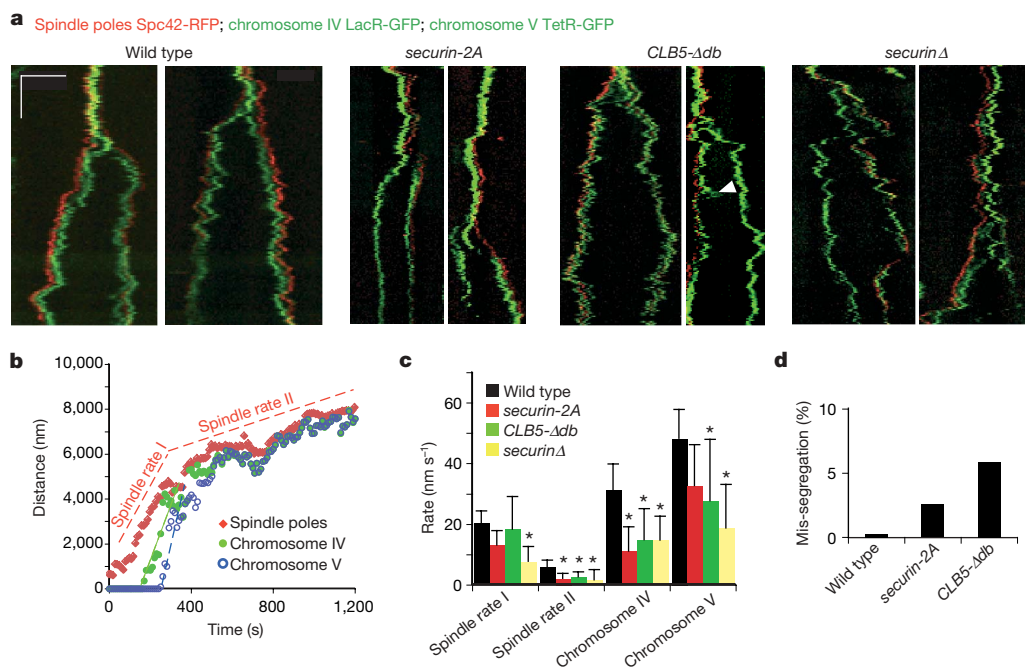
We next analysed anaphase in the *securin-2A* strain, in which we predict the normal delay of separase activation is lost, resulting in cohesin cleavage early relative to dephosphorylation of Cdk1 substrates. We also analysed the effects of the *CLB5-Δdb* mutation, which is likewise predicted to introduce a delay in the dephosphorylation of Cdk1 substrates relative to separase activation. Finally, we analysed *securinΔ* cells, where separase activation is less tightly linked to changes in Cdk1 activity. We observed a wide range of anaphase defects in all mutants (Fig. 4). In a subset of *securin-2A* cells, the disjunction of chromatids was not coordinated with elongation of the spindle, chromosomes failed to segregate processively and the rates of chromatid separation at anaphase A were significantly reduced (Fig. 4a, c). Spindle defects were more pronounced in *CLB5-Δdb* and *securinΔ* cells, in which the spindle frequently failed to elongate or elongated and then collapsed back to a shorter length (Fig. 4a). Our results are consistent with a general role for securin in the coordination of chromosome disjunction with changes in the

phosphorylation state of Cdk1 substrates on the spindle and kinetochores. They also support the specific hypothesis that modulation of securin ubiquitination by phosphorylation helps coordinate the events of anaphase.

We assessed the accuracy of anaphase in our GFP-labelled yeast strain. Each cell in this strain is expected to have two GFP dots before anaphase, and cells containing more than two dots are products of a mis-segregation event. We found that loss of phosphoregulation of securin in the *securin-2A* mutant led to an elevated rate of chromosome mis-segregation, as did inappropriate Cdk1 activity in the *CLB5-Δdb* mutant (Fig. 4d).

We conclude that positive feedback in securin regulation promotes accurate chromosome segregation. Positive feedback could generate bistability in anaphase control<sup>17,18</sup>, ensuring that the cell is irreversibly committed to complete sister-chromatid separation. Synchronous sister separation may have several benefits: for example, if disjunction of the first chromatid pair is sensed by the spindle assembly checkpoint, then rapid separation of the remaining sisters would reduce the danger of inappropriate checkpoint activation before anaphase is complete. In addition, the initiation of chromosome segregation might cause local perturbations to the spindle, making it important to transition rapidly to the segregation mode of spindle function to avoid mechanical catastrophe.

Our studies support the possibility that each sister chromatid pair separates at a distinct threshold of separase activity (Supplementary Fig. 9). Such differences in separase sensitivity would help explain the observation that sister chromatids separate in a specific order, as we observed in yeast and as previously observed in plant and animal cells<sup>26,27</sup>. The ordered segregation of chromatids provides a mechanism to maintain nuclear architecture through successive generations<sup>27</sup>. However, the existence of distinct separase thresholds provides a challenge for the rapid execution of anaphase. The most



**Figure 4 | Modulation of securin ubiquitination by Cdk1 helps coordinate anaphase onset with changes in spindle dynamics.** **a**, Kymographs from wild-type and mutant cells. The x axis is a one-dimensional projection along the spindle axis (scale bar, 5  $\mu$ m) and the y axis is time (scale bar, 5 min; total time, 30 min). **b**, Spindle length (red diamonds) and chromosome segregation (green dots, chromosome IV; blue circles, chromosome V) in a wild-type anaphase, using a spot-tracking algorithm<sup>30</sup>. Rates of the two phases of spindle elongation are highlighted by red dashed lines; initial rates of chromatid segregation are highlighted with green and blue dashed lines.

When the daughter spindle pole was difficult to detect because of the slow maturation of the dsRed fluorophore, the position of the associated sister chromatid was used as a proxy for spindle-pole position after anaphase A. **c**, Average rates of spindle elongation and sister-chromatid separation in wild-type and mutant cells. Error bars, s.d. ( $n = 10$  cells). Asterisks,  $P < 0.05$  ( $t$ -test) when comparing the wild type with the mutant. **d**, Rate of chromosome mis-segregation in wild-type and mutant cells. Only cells with more than the expected number of GFP dots were scored as a mis-segregation event.



effective mechanism for achieving chromosome segregation that is both ordered and abrupt is to make separase activation rapid and switch-like.

## METHODS SUMMARY

**Yeast strains.** *Saccharomyces cerevisiae* strains were derivatives of a W303 diploid strain containing: *SPC42-dsRed::NAT*; *trp1::256×LACO:TRP1/trp1-1*; *his3::CUP1-GFP-LACI:His3/his3-11*; *ura3::112×TETO:URA3/ura3-1*; *leu2::TETR-GFP:LEU2/leu2-1*. This strain was created by mating strain LH557 (*Mat a*; *trp1::256×LACO:TRP1*; *his3::CUP1-GFP-LACI:His3*; *SPC42-dsRed-NAT*) with LH554 (*Mat alpha*; *leu2::TETR-GFP:LEU2*; *ura3::112×TETO:URA3*; *SPC42-dsRed-NAT*).

**Protein purification.** Securin-3FLAG-6HIS was purified by the tandem affinity method<sup>9</sup>. APC, Ubc4, Uba1 and Cdh1 were purified as described<sup>28</sup>. Clb2 was expressed from a 2  $\mu$  *P<sub>GALI</sub>*-TAP plasmid, and Clb2-Cdk1 complexes were purified using a C-terminal TAP tag on Clb2 (ref. 29). GST-Cdc14 was purified from bacteria<sup>15</sup>. Securin and Cdc20 were expressed in the TNT reticulocyte lysate-coupled transcription/translation system (Promega) with a C- and N-terminal TEV-ZZ tag, respectively, purified rapidly on M-270 epoxy dynabeads (Invitrogen) coupled to IgG, and eluted with TEV protease.

**Mass spectrometry.** Identification of phosphorylation sites was performed using a modular mass spectrometric tool as described<sup>9</sup>.

**Ubiquitination assays.** Ubiquitination assays were performed as described<sup>28</sup>, except where securin was pre-incubated with Cdk1-Clb2 in a 10- $\mu$ l reaction volume containing 25 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol, 0.1 mg ml<sup>-1</sup> BSA, 1 mM ATP, and 10 mM MgCl<sub>2</sub> for 20 min at room temperature. Kinase reactions were terminated by incubation for 5 min with 50 mM EDTA. In some cases, Cdc14 was then added to this mixture for a further 20 min at room temperature before ubiquitination assays.

**Microscopy.** All microscopy was performed in the University of California, San Francisco, Nikon Imaging Center using a TE2000U inverted microscope (Nikon) with Yokogawa CSU22 spinning disk confocal illumination (Solamere Technology Group) and a Cascade II CCD Camera (Photometrics). Stacks of seven images 1  $\mu$ m apart were acquired every 10 s for 30 min at 25 °C. Images were acquired using micro-manager software (<http://micro-manager.org/>) and analysed using ImageJ (<http://rsb.info.nih.gov/ij/>) with the SpotTracker2D plugin (<http://bigwww.epfl.ch/sage/soft/spottracker/gasser.html>).

Received 5 February; accepted 24 April 2008.

Published online 15 June 2008.

- Nasmyth, K. Segregating sister genomes: the molecular biology of chromosome separation. *Science* **297**, 559–565 (2002).
- Peters, J. M. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Rev. Mol. Cell Biol.* **7**, 644–656 (2006).
- Thornton, B. R. & Toczyski, D. P. Precise destruction: an emerging picture of the APC. *Genes Dev.* **20**, 3069–3078 (2006).
- Sullivan, M. & Morgan, D. O. Finishing mitosis, one step at a time. *Nature Rev. Mol. Cell Biol.* **8**, 894–903 (2007).
- Stegmeier, F., Visintin, R. & Amon, A. Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell* **108**, 207–220 (2002).
- Queralt, E., Lehane, C., Novak, B. & Uhlmann, F. Downregulation of PP2A(Cdc55) phosphatase by separase initiates mitotic exit in budding yeast. *Cell* **125**, 719–732 (2006).
- Agarwal, R., Tang, Z., Yu, H. & Cohen-Fix, O. Two distinct pathways for inhibiting pds1 ubiquitination in response to DNA damage. *J. Biol. Chem.* **278**, 45027–45033 (2003).
- Wang, H. *et al.* Pds1 phosphorylation in response to DNA damage is essential for its DNA damage checkpoint function. *Genes Dev.* **15**, 1361–1372 (2001).
- Blethrow, J. D., Tang, C., Deng, C. & Krutchinsky, A. N. Modular mass spectrometric tool for analysis of composition and phosphorylation of protein complexes. *PLoS ONE* **2**, e358 (2007).
- Hornig, N. C., Knowles, P. P., McDonald, N. Q. & Uhlmann, F. The dual mechanism of separase regulation by securin. *Curr. Biol.* **12**, 973–982 (2002).

- Agarwal, R. & Cohen-Fix, O. Phosphorylation of the mitotic regulator Pds1/securin by Cdc28 is required for efficient nuclear localization of Esp1/separase. *Genes Dev.* **16**, 1371–1382 (2002).
- King, R. W., Glotzer, M. & Kirschner, M. W. Mutagenic analysis of the destruction signal of mitotic cyclins and structural characterization of ubiquitinated intermediates. *Mol. Biol. Cell* **7**, 1343–1357 (1996).
- Wan, J., Xu, H. & Grunstein, M. CDC14 of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **267**, 11274–11280 (1992).
- Visintin, R. *et al.* The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol. Cell* **2**, 709–718 (1998).
- Jaspersen, S. L., Charles, J. F. & Morgan, D. O. Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Curr. Biol.* **9**, 227–236 (1999).
- Stegmeier, F. & Amon, A. Closing mitosis: the functions of the Cdc14 phosphatase and its regulation. *Annu. Rev. Genet.* **38**, 203–232 (2004).
- Ferrell, J. E. Jr. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14**, 140–148 (2002).
- Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).
- Straight, A. F., Marshall, W. F., Sedat, J. W. & Murray, A. W. Mitosis in living budding yeast: anaphase A but no metaphase plate. *Science* **277**, 574–578 (1997).
- Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35–45 (1997).
- Wäsch, R. & Cross, F. APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. *Nature* **418**, 556–562 (2002).
- Higuchi, T. & Uhlmann, F. Stabilization of microtubule dynamics at anaphase onset promotes chromosome segregation. *Nature* **433**, 171–176 (2005).
- Pereira, G. & Schiebel, E. Separase regulates INCENP-Aurora B anaphase spindle function through Cdc14. *Science* **302**, 2120–2124 (2003).
- Woodbury, E. L. & Morgan, D. O. Cdk and APC activities limit the spindle-stabilizing function of Fin1 to anaphase. *Nature Cell Biol.* **9**, 106–112 (2007).
- Parry, D. H., Hickson, G. R. & O'Farrell, P. H. Cyclin B destruction triggers changes in kinetochore behavior essential for successful anaphase. *Curr. Biol.* **13**, 647–653 (2003).
- Vig, B. K. Sequence of centromere separation: occurrence, possible significance, and control. *Cancer Genet. Cytogenet.* **8**, 249–274 (1983).
- Gerlich, D. *et al.* Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* **112**, 751–764 (2003).
- Carroll, C. W. & Morgan, D. O. Enzymology of the anaphase-promoting complex. *Methods Enzymol.* **398**, 219–230 (2005).
- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Sage, D. *et al.* Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE Trans. Image Process.* **14**, 1372–1383 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank: P. H. O'Farrell, A. D. Johnson and M. J. Sullivan for discussions; J. A. Ubersax, G. Goshima and O. Cohen-Fix for reagents; S. Foster, M. C. Rodrigo-Brenni, M. Enquist-Newman and the Morgan laboratory for help generating strains and reagents; J. M. Pedraza and A. van Oudenaarden for help with the model; K. S. Thorn and the University of California, San Francisco, Nikon Imaging Center for help with microscopy; and M. J. Sullivan and J. L. Feldman for reading the manuscript. This work was supported by funding from the National Institute of General Medical Sciences (D.O.M.), a grant from the Sandler Family Foundation (A.N.K.) and a fellowship from the National Science Foundation (L.J.H.).

**Author Contributions** L.J.H. designed, performed and analysed the experiments; A.N.K. performed mass spectrometric analysis; D.O.M. provided guidance.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.O.M. ([david.morgan@ucsf.edu](mailto:david.morgan@ucsf.edu)).

## LETTERS

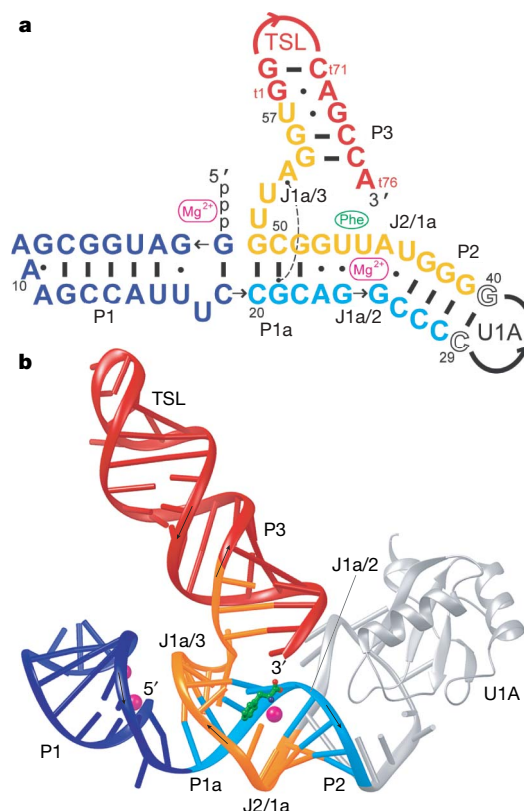
Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozymeHong Xiao<sup>1</sup>, Hiroshi Murakami<sup>2</sup>, Hiroaki Suga<sup>2,3</sup> & Adrian R. Ferré-D'Amaré<sup>1</sup>

In modern organisms, protein enzymes are solely responsible for the aminoacylation of transfer RNA. However, the evolution of protein synthesis in the RNA world required RNAs capable of catalysing this reaction. Ribozymes that aminoacylate RNA by using activated amino acids have been discovered through selection *in vitro*<sup>1–5</sup>. Flexizyme is a 45-nucleotide ribozyme capable of charging tRNA *in trans* with various activated L-phenylalanine derivatives. In addition to a more than 10<sup>5</sup> rate enhancement and more than 10<sup>4</sup>-fold discrimination against some non-cognate amino acids, this ribozyme achieves good regioselectivity: of all the hydroxyl groups of a tRNA, it exclusively aminoacylates the terminal 3'-OH<sup>5–7</sup>. Here we report the 2.8-Å resolution structure of flexizyme fused to a substrate RNA. Together with randomization of ribozyme core residues and reselection, this structure shows that very few nucleotides are needed for the aminoacylation of specific tRNAs. Although it primarily recognizes tRNA through base-pairing with the CCA terminus of the tRNA molecule, flexizyme makes numerous local interactions to position the acceptor end of tRNA precisely. A comparison of two crystallographically independent flexizyme conformations, only one of which appears capable of binding activated phenylalanine, suggests that this ribozyme may achieve enhanced specificity by coupling active-site folding to tRNA docking. Such a mechanism would be reminiscent of the mutually induced fit of tRNA and protein employed by some aminoacyl-tRNA synthetases<sup>8,9</sup> to increase specificity.

Flexizyme was selected as a *cis*-acting 5'-leader sequence of tRNA that catalyses the aminoacylation of the 3'-OH of tRNA employing activated L-phenylalanine derivatives such as L-phenylalanyl-cyanomethyl ester and L-phenylalanyl-adenylate<sup>5</sup>. Flexizyme can be separated from its attached tRNA by RNase P. The released 45-nucleotide ribozyme is active *in trans*, with an efficiency comparable to that shown *in cis*<sup>7</sup>. Previous analyses indicate that, like some protein aminoacyl-tRNA synthetases (ARSs)<sup>10</sup>, flexizyme recognizes solely the tRNA acceptor stem and is fully active with minihelices (comprising the tRNA acceptor stem and T stem-loop)<sup>11</sup>. To understand how this compact ribozyme achieves high regioselectivity, we solved the crystal structure of a flexizyme–tRNA minihelix fusion. This was facilitated by replacing a functionally dispensable loop<sup>7</sup> of flexizyme with the binding site for the U1A protein. Crystallization of the reaction product (L-phenylalanyl-RNA) was precluded by its hydrolytic instability. Instead, L-phenylalanyl-ethyl ester (PheEE), a mimic of L-phenylalanyl-cyanomethyl ester, was soaked into crystals (Methods).

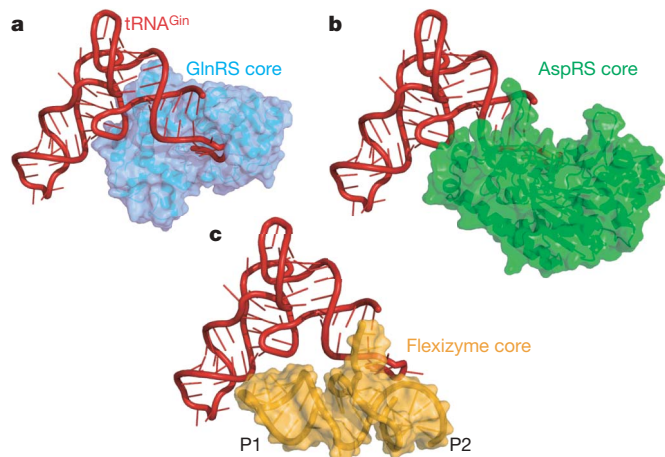
The flexizyme core consists of a coaxial stack of four helices, one irregular and three A-form (paired regions P1, P1a and P2), and a 3' extension (Fig. 1). The irregular helix comprises two segments (joining regions J1a/2 and J2/1a) that have previously been shown biochemically to participate in amino-acid recognition but were thought to be single-stranded<sup>12</sup> (Supplementary Fig. 1). Residues 52–54 near the 3' end of the

ribozyme form a hairpin-shaped turn (J1a/3), helping project the last three residues (55–57) of flexizyme away from the ribozyme helical stack. These nucleotides form helix P3 by base-pairing with the complementary residues tG73–tC75 of the minihelix ('t' denotes tRNA residues), juxtaposing tA76 (the site of aminoacylation) with the broad minor groove of the irregular helix. Flexizyme contacts the minihelix only on its acceptor end, and our structure is compatible with full-length tRNA binding to the ribozyme in the same orientation (Supplementary Fig. 2). Natural ribozymes, including the ribosome<sup>13</sup> and RNase P (ref. 14), also recognize tRNA by base-pairing with the CCA terminus.



**Figure 1 | Overall structure of the flexizyme–tRNA minihelix fusion.** **a**, Sequence and secondary structure of the crystallization construct. Binding sites of well-ordered Mg<sup>2+</sup> ions and of phenylalanine are indicated. The dashed line denotes an A-minor interaction. The minihelix (red) is numbered by using the tRNA convention ('t' precedes residues). T stem-loop (TSL) and U1A-binding loop are described in Methods. **b**, Diagram of the three-dimensional structure. Phenylalanine is depicted in green, and Mg<sup>2+</sup> ions in magenta.

<sup>1</sup>Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024, USA. <sup>2</sup>Research Center for Advanced Science and Technology, The University of Tokyo, 153-8904 Tokyo, Japan. <sup>3</sup>Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 113-8656, Tokyo, Japan.

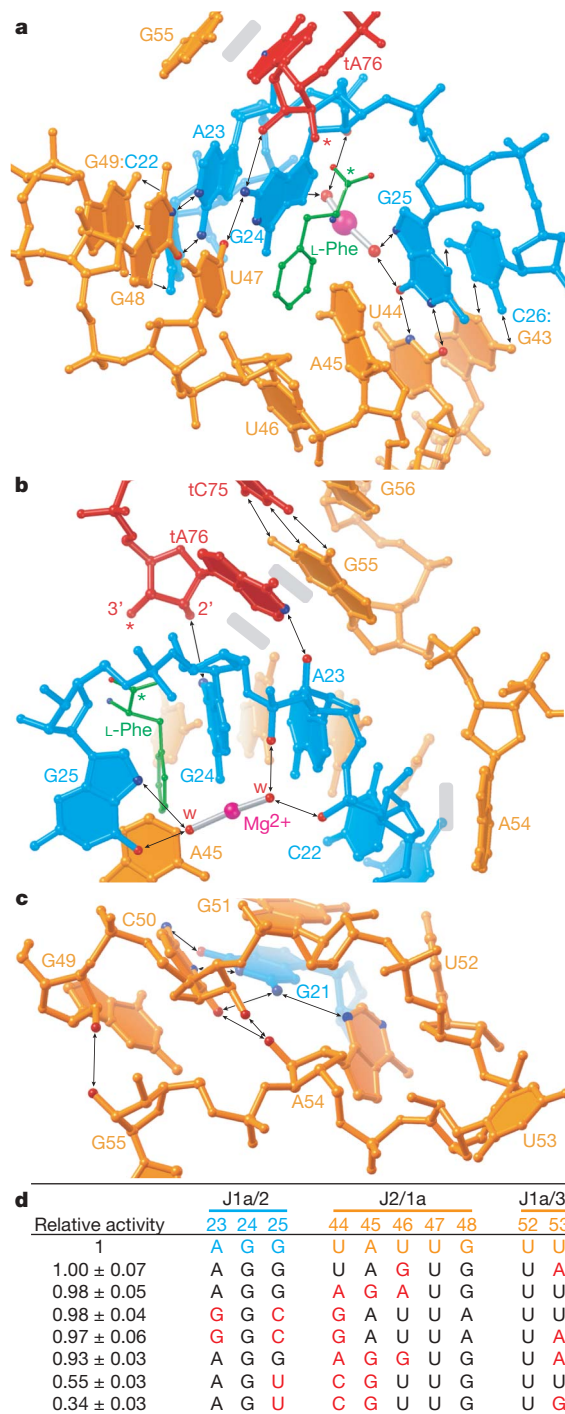


**Figure 2 | Comparison of protein and RNA aminoacyl-tRNA synthetases.** **a**, Core of *Escherichia coli* glutamyl-tRNA synthetase (GlnRS; cyan), a class I ARS, bound to tRNA<sup>Gln</sup> (ref. 16). **b**, Core of *E. coli* aspartyl-tRNA synthetase (AspRS; green), a class II ARS, bound to tRNA<sup>Gln</sup><sup>17</sup>. Its cognate tRNA was superimposed on tRNA<sup>Gln</sup>, and the latter is shown. **c**, Flexizyme core (omitting the U1A-binding loop) docked onto tRNA<sup>Gln</sup> by superimposing the minihelix on the tRNA acceptor stem.

Protein ARSs belong to two structurally distinct families<sup>15</sup>. Class I enzymes approach the acceptor stem from the minor groove and initially aminoacylate the terminal 2'-OH of tRNA. Class II enzymes approach from the major groove and initially aminoacylate the 3'-OH (Fig. 2a, b). After charging, the acyl group equilibrates between the 2'-OH and the 3'-OH. Superposition of flexizyme on representative class I and II ARS-tRNA complexes shows that flexizyme approaches tRNA in a manner analogous to that of class II ARSs (Fig. 2c). It was shown previously<sup>6</sup> that flexizyme aminoacylates tRNA analogues lacking the terminal 2'-OH group 2.4-fold more slowly than it does wild-type tRNA. In contrast, the ribozyme showed a 200-fold decrease in activity when presented with a tRNA analogue lacking a 3'-OH. Thus, flexizyme preferentially aminoacylates the tRNA 3'-OH. Mimicry of class II ARSs by flexizyme seems to result from both a common direction of approach to the tRNA substrate, and the conformation of the enzyme-bound CCA terminus. Unlike class I ARSs, which impose a kink on this single-stranded segment of tRNA<sup>16</sup>, class II ARSs bind to a CCA in a near-helical conformation<sup>17</sup>. Flexizyme base-pairs to the CCA, imposing a helical conformation on it. In double-helical RNA, the 2'-OH is buried in the minor groove, making it less accessible to an enzyme than the 3'-OH.

The irregular helix harbouring the active site of flexizyme is formed by three non-canonical base pairs (A23•G48, G24•U47 and G25•U44) and two unpaired nucleotides (A45 and U46; Fig. 3a). The pair formed between the Watson-Crick faces of A23 and G48 widens the minor groove, allowing only the O4 carbonyl oxygen of U47 to interact with G24. Juxtaposition of these two 'stretched' non-canonical pairs and the two unpaired nucleotides results in the amino-acid-binding pocket. A hydrated Mg<sup>2+</sup> ion in the major groove of the active-site helix (Fig. 3b) bridges the Hoogsteen face of G25 and the backbone phosphates of A23 and G24, stabilizing the active site.

Flexizyme precisely positions the acceptor end of its substrate tRNA by four methods. First, the J1a/3 hairpin (Fig. 3b, c) helps to place G55, which pairs with tC75 to form the bottom base pair of P3, perpendicular to the main helical stack of the ribozyme. A54 of J1a/3 packs against the C50•G21 base pair of P1a through a canonical class I A-minor interaction<sup>18</sup>. U52 stacks on A54 and also packs against P1a, and U53 is extruded from the turn. Mutation of A54 to U abrogates 95% of ribozyme activity<sup>12</sup>, suggesting that the A-minor interaction is required for activity. Second, in addition to base-pairing with tC75, G55 makes a partial cross-strand stack with the base of tA76 (Fig. 3a, b). Replacement of G55•tC75 with the corresponding C55•tG75



**Figure 3 | Structure and sequence requirements of the active site.** **a**, View into the minor groove. The ethyl group of the bound L-phenylalanine-ethyl ester, lacking electron density, is omitted. Asterisks denote the two reactive atoms; grey rectangles indicate some stacking interactions. Tb<sup>3+</sup> cleavage assays<sup>30</sup> indicate the presence, close to J2/1a, of an outer-sphere coordinated divalent cation required for activity. This cation may correspond to the Mg<sup>2+</sup> ion shown. Two inner-sphere coordination water molecules are resolved crystallographically. **b**, View from the major groove. The Mg<sup>2+</sup> ion lies about 8 Å away from the reactive carbonyl; its electrostatic contribution to transition-state stabilization is probably modest. **c**, The J1a/3 hairpin turn. **d**, Summary of a of reselection experiment in which the ten listed residues were randomized. The activity of reselected sequences (mean ± s.d. for three independent measurements) is normalized to that of the parental sequence. Nucleotides that varied between the reselected and parental sequences are in red, and those that did not in black.



mutations decreases flexizyme activity by 90% (ref. 12). Because G forms the most stable stacking interactions<sup>19</sup>, our structure suggests that its partial cross-strand stacking on tA76 is needed for flexizyme function. Third, the base of tA76 makes van der Waals contact with the ribose of G24, and a hydrogen bond between its N1 and the 2'-OH of A23 (Fig. 3b). Mutation of tA76 to G or U results in 85% and 90% loss of activity, respectively<sup>12</sup>, suggesting that the specific interactions made by the adenine are important for activity. Fourth, the 2'-OH of tA76 hydrogen bonds with N2 of G24 (Fig. 3a). None of these interactions require a covalent bond between the two RNAs. Because the activities of *cis*- and *trans*-flexizymes are comparable<sup>7</sup>, the structure of the *trans*-flexizyme-tRNA complex is probably similar to what we observe.

Although weak, which is consistent with the high  $K_m$  of flexizyme for activated phenylalanine (more than 5 mM)<sup>7</sup>, residual  $|F_o| - |F_c|$  electron density for the benzyl group of the inhibitor PheEE is present in the pocket formed by juxtaposition of the U47•G24 pair with the unstacked U46 in the active site of one of the two crystallographically independent conformers of flexizyme in our crystals (Fig. 3a and Supplementary Fig. 3). The phenyl group stacks on G24, maximizing the favourable interaction<sup>20</sup> of the O6 of the purine with the partial positive charge at the centre of the phenyl ring (Fig. 3b). The importance of this oxygen–aromatic interaction is supported by the marked preference of flexizyme for aromatic substrates (Supplementary Figs 4 and 5). An adjacent depression formed between the splayed bases of G24 and G25 accommodates the  $\alpha$ -carbon and the carbonyl group of the amino acid (Fig. 3a). No electron density is observed for the ethyl group of PheEE, but space is present for it to project away from the RNA, which is consistent with the tolerance of flexizyme for bulky

leaving groups, including AMP<sup>5</sup>. The amine of the phenylalanine points directly out of the groove, which is consistent with the lack of recognition of this functional group by flexizyme<sup>5</sup>.

To verify the functional importance of flexizyme core nucleotides, we reselected seven new active variants (Fig. 3d) from a pool of flexizyme–microhelix RNAs containing random sequences in the active site and the hairpin turn<sup>21</sup>. The reselection indicates that the identities of only two nucleotides in the irregular active site helix, G24 and U47, are essential. The unstacked U46 can be replaced with a purine, and position 45 can be either purine. The expansion of the minor groove produced by the A23•G48 pair is important, because it or the reverse G23•A48 pair is conserved. U52 from the apex of the hairpin turn is also absolutely conserved. This underscores the importance of J1a/3 for activity.

In the two flexizyme conformers in our structure, J1a/3 functions as a hinge between the ribozyme helical stack and the minihelix (Fig. 4a, b). In molecule B, the J1a/3 hairpin is not fully docked with P1a (Fig. 4c), the 2'-OH of tA76 does not engage the base of G24, and the functionally critical G24•U47 pair of the active site is not formed (Fig. 4d and Supplementary Fig. 6). Also in molecule B, U47 buckles towards U46, obstructing the binding of PheEE. On the basis of our crystallographic and reselection studies, we speculate that adoption by the active site of flexizyme of a conformation capable of binding and positioning activated phenylalanine is intimately coupled to the docking of J1a/3 and P1a. This coupling of tRNA binding and active-site folding would be reminiscent of the indirect readout of acceptor stem conformation by some protein ARSs<sup>22</sup>.

To function in translation, tRNA needs to be aminoacylated exclusively at its 3'-terminal residue. This regioselective aminoacylation may have evolved in the RNA world to recruit the replicase ribozyme to the tRNA-like ends of the genome, thereby promoting full-length replication<sup>23,24</sup>. Alternatively, it may have evolved to expand the chemical diversity of primitive ribozyme active sites, in an analogous manner to the prosthetic groups of proteins<sup>25</sup>. Either model requires ARS ribozymes capable of regioselective aminoacylation. Together with previous analyses, our structures and reselection show that remarkably few active-site and J1a/3 residues are needed for flexizyme activity. The simplicity of flexizyme supports the hypothesis that ribozymes were responsible for all biochemical catalysis before the evolution of translation. It is possible that this small ribozyme achieves regioselectivity through an induced-fit mechanism. Protein enzymes, including ARSs<sup>8,9</sup>, frequently attain enhanced specificity by coupling active-site folding to substrate binding, but previous structural studies of catalytic RNAs have mostly revealed rigid, pre-organized small-molecule binding sites<sup>26–28</sup>. Our demonstration that an ARS ribozyme evolved *in vitro* achieves specificity by interacting only with the acceptor stem of tRNA is consistent with the proposal<sup>29</sup> that class II ARS that are active with tRNA minihelices are closely related evolutionarily to the primordial ARSs.

## METHODS SUMMARY

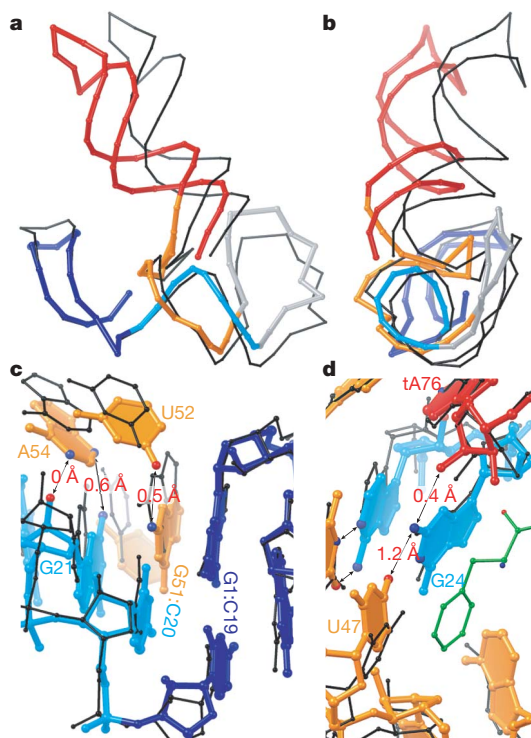
The flexizyme–minihelix fusion RNA was cleaved with the VS ribozyme to generate a homogeneous 3' end. The 2',3'-cyclic phosphate was opened and removed with T4 polynucleotide kinase (PNK) (Supplementary Fig. 7). The RNA was co-crystallized with selenomethionyl U1A protein and the structure was determined by a combination of multiwavelength anomalous diffraction (MAD) and molecular replacement (Supplementary Fig. 8). A second structure was solved with amplitudes from a crystal soaked in a cryoprotectant including 2.5 mM PheEE (the solubility limit of the compound). The models have been refined to  $R_{\text{free}}/R_{\text{work}}$  of 27.9%/22.3% and 30.4%/25.4% against all data extending to 2.8 Å and 3.0 Å, respectively (Supplementary Table 1).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 January; accepted 28 April 2008.

Published online 11 June 2008.

1. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).



**Figure 4 | Conjectural coupling of tRNA docking to active-site folding.** **a**, Superposition of the two crystallographically independent flexizyme structures. Throughout, molecule B (undocked conformation) is in black. **b**, Orthogonal view. **c**, Comparison of the interface between J1a/3 and P1a in the two molecules. P1 adopts similar conformations in both molecules; J1a/3 residues (U52 and A54) are farther from P1a in molecule B. Red numbers denote the increase in length of the indicated hydrogen bonds from molecule A to molecule B. Differences are mostly comparable to the precision of the structures (Methods); however, distances are systematically larger for molecule B. **d**, Buckling of U47 into the phenylalanine-binding site in molecule B obstructs the binding of the amino acid.

2. Illangasekare, M. & Yarus, M. A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* **5**, 1482–1489 (1999).
3. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl Acad. Sci. USA* **96**, 5470–5475 (1999).
4. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nature Struct. Biol.* **7**, 28–33 (2000).
5. Saito, H., Kourouklis, D. & Suga, H. An *in vitro* evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797–1806 (2001).
6. Saito, H. & Suga, H. A ribozyme exclusively aminoacylates the 3'-hydroxyl group of the tRNA terminal adenosine. *J. Am. Chem. Soc.* **123**, 7178–7179 (2001).
7. Murakami, H., Saito, H. & Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **10**, 655–662 (2003).
8. Cusack, S., Yaremchuk, A. & Tkalco, M. The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with tRNA<sup>Ser</sup> and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J.* **15**, 2834–2842 (1996).
9. Moulinier, L. *et al.* The structure of an AspRS-tRNA<sup>Asp</sup> complex reveals a tRNA-dependent control mechanism. *EMBO J.* **20**, 5290–5301 (2001).
10. Schimmel, P., Giegé, R., Moras, D. & Yokoyama, S. An operational code for amino acids and possible relationship to genetic code. *Proc. Natl Acad. Sci. USA* **90**, 8763–8768 (1993).
11. Ramaswamy, K., Saito, H., Murakami, H., Shiba, K. & Suga, H. Designer ribozymes: programming the tRNA specificity into flexizyme. *J. Am. Chem. Soc.* **126**, 11454–11455 (2004).
12. Saito, H., Watanabe, K. & Suga, H. Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA* **7**, 1867–1878 (2001).
13. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
14. Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13**, 4870–4876 (1994).
15. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206 (1990).
16. Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. Structure of *E. coli* glutamyl-tRNA synthetase complexed with tRNA<sup>Gln</sup> at 2.8 Å resolution: implications for tRNA discrimination. *Science* **246**, 1135–1142 (1989).
17. Ruff, M. *et al.* Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA<sup>Asp</sup>. *Science* **252**, 1682–1689 (1991).
18. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA* **98**, 4899–4903 (2001).
19. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, New York, 1984).
20. Burley, S. K. & Petsko, G. A. Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125–192 (1988).
21. Murakami, H., Ohta, A., Ashiai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3**, 357–359 (2006).
22. Perona, J. J. & Hou, Y. M. Indirect readout of tRNA for aminoacylation. *Biochemistry* **46**, 10419–10432 (2007).
23. Weiner, A. M. & Maizels, N. tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl Acad. Sci. USA* **84**, 7383–7387 (1987).
24. Orgel, L. E. The origin of polynucleotide-directed protein synthesis. *J. Mol. Evol.* **29**, 465–474 (1989).
25. Wong, J. T. Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Orig. Life Evol. Biosph.* **21**, 165–176 (1991).
26. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**, 259–264 (1998).
27. Serganov, A. *et al.* Structural basis for Diels–Alder ribozyme catalyzed carbon–carbon bond formation. *Nature Struct. Mol. Biol.* **12**, 218–224 (2005).
28. Klein, D. J. & Ferré-D'Amaré, A. R. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756 (2006).
29. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–5035 (1998).
30. Saito, H. & Suga, H. Outersphere and innersphere coordinated metal ions in an aminoacyl-tRNA synthetase ribozyme. *Nucleic Acids Res.* **30**, 5151–5159 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the staff at ALS beamline 5.0.2 and J. Bolduc for assistance with synchrotron and in-house X-ray data collection, respectively, and T. Edwards, T. Hamma, D. Klein, J. Pitt, J. Posakony, A. Roll-Mecak and B. Shen for discussions. A.R.F. is a Distinguished Young Scholar in Medical Research of the W. M. Keck Foundation. This work was supported by grants from research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (to H.S.) and the W. M. Keck Foundation (to A.R.F.).

**Author Information** Atomic coordinates and structure factor amplitudes for aminoacyl-tRNA synthetase ribozyme–minihelix fusions refined against crystal II and crystal III data have been deposited with the Protein Data Bank with accession codes 3CUL and 3CUN, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.R.F. (aferre@fhcrc.org).

## METHODS

**Protein and RNA preparation.** Expression of the selenomethionyl U1A RNA-binding domain (RBD) double mutant (Y31H, Q36R) was as described<sup>31</sup>. Plasmid pMU40 is a pUC19 derivative encoding the flexizyme–minihelix fusion used for structure determination followed by the substrate stem-loop of the VS ribozyme<sup>32</sup>. The TSL (residues t3–t70 in tRNA numbering; Fig. 1a) is 5'-UGGUACGAGGUUCGAAUCCUCGUACCG-3'. The U1A-binding loop (residues 30–39) is 5'-AUUGCACUCC-3'. *In vitro* transcription reactions (310 K, 4 h) contained 75 mg l<sup>-1</sup> BamHI-linearized pMU40, 50 mg l<sup>-1</sup> SspI-linearized pAVA (ref. 32), each NTP at 2.5 mM, 30 mM Tris-HCl pH 8.1, 40 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 2 mM spermidine, 10 mM dithiothreitol, 1,000 U l<sup>-1</sup> *E. coli* inorganic pyrophosphatase<sup>33</sup> and 0.1 g l<sup>-1</sup> T7 RNA polymerase. The VS ribozyme-cleaved transcript<sup>32</sup> was purified by denaturing PAGE, passively eluted into water, and desalted by ultrafiltration. The 2',3'-cyclic phosphate was removed by treatment of the RNA with PNK (New England Biolabs). Reactions (310 K, 5 h) contained 68 µM RNA, PNK (1 U of enzyme per 0.14 nmol of RNA), 25 mg l<sup>-1</sup> BSA, 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol and 1 mM ATP (Supplementary Fig. 7). After extraction with phenol/chloroform and precipitation with ethanol, the RNA was desalted by ultrafiltration, concentrated to about 0.9 mM and stored at 277 K.

**Crystallization and diffraction data collection.** RNA (0.74 mM in 10 mM Hepes-KOH pH 7.5, 1 mM EDTA) was snap-cooled by heating to 363 K for 2 min and transferring to ice-water for 5 min. A 1.25-fold molar excess of selenomethionyl U1A-RBD double mutant was added to the RNA, and the solution was adjusted to 10 mM MgCl<sub>2</sub>, 1 mM cobalt(III) hexammine, 0.6 mM PheEE (Sigma), 1 mM spermine, and the RNA at 0.3 mM. Sitting drops prepared by mixing 1 µl of this solution with 1 µl of a reservoir solution comprising 100 mM magnesium formate pH 7.0, 14.5–15% (w/v) PEG 3000 and 0.8–1.0 M LiCl were equilibrated by vapour diffusion at 295 K. Crystals with the symmetry of space group C2 (unit cell parameters are given in Supplementary Table 1) appeared in days and reached maximum dimensions of 0.3 × 0.2 × 0.14 mm<sup>3</sup> over weeks. Crystals were transferred over 15 min to a solution of the same composition as the reservoir solution, but with the concentration of LiCl increased to 1.8 M, and flash-frozen by plunging into liquid nitrogen. For crystal III, the concentration of PheEE in the cryoprotectant was increased to 2.5 mM (this was limited by poor solubility of PheEE). Diffraction data (Supplementary Table 1) were collected at 100 K (with the inverse-beam method for crystal I) at beamline 5.0.2 of the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, and reduced with the HKL package<sup>34</sup>.

**Structure determination and refinement.** Six selenium sites were located and their parameters were refined, and two energy MAD phases were calculated with CNS<sup>35</sup> using crystal I data (Supplementary Table 1). The electron-density map resulting from density modification of these phases clearly showed the U1A-RBD and its binding site but did not reliably indicate the path of the RNA chain. The program PHASER<sup>36</sup> was used for molecular replacement with the use of crystal II data and two copies each (six search models) of the U1A-RBD and 15 nucleotides of its binding site, an 8-nucleotide stem-loop comprising a canonical GAAA tetraloop, and a 31-nucleotide stem-loop comprising a canonical tRNA minihelix. The six strongest peaks from an anomalous difference Fourier synthesis calculated with amplitudes from the peak data set of crystal I and model phases from the top PHASER solution coincided with the selenomethionines in the U1A-RBD models of the PHASER solution. Phase combination of the MAD phases and the PHASER model phases produced electron-density maps

(Supplementary Fig. 8) that allowed unambiguous tracing of the RNA through iterative rounds of model building<sup>37</sup>, simulated annealing, energy minimization and restrained individual isotropic *B*-factor refinement<sup>35</sup>. A maximum-likelihood target was employed for refinement against all structure-factor amplitudes from crystal II, and experimental phase-probability distributions. A solvent mask and an anisotropic *B*-factor correction were used throughout. The solvent mask parameters (solvent density 0.3; solvent *B*-factor 25 Å<sup>2</sup>) were chosen empirically to minimize the *R*<sub>free</sub> of the lower-resolution shells. The final crystallographic model (Supplementary Table 1) comprises two complete RNA molecules (92 residues each), two selenomethionyl U1A-RBDs (residues 1–92 and 1–91), ten Mg<sup>2+</sup> ions, one K<sup>+</sup> ion and ten water molecules (5,346 non-hydrogen atoms). Ramachandran analysis shows that 86.8% and 12.6% of the protein residues have most favoured and additionally allowed backbone conformations, respectively; there are no residues with a disallowed conformation. All figures, except Fig. 4, are of the docked form of the RNA (molecule A) and were prepared with PyMol<sup>38</sup> or RIBBONS<sup>39</sup>. Refinement against crystal III amplitudes started with the model from crystal II refinement. Waters and ions were stripped from the model, and after iterative rounds of simulated annealing, energy minimization and tightly restrained individual isotropic *B*-factor refinement, ions were placed into the strongest features of the residual  $|F_o| - |F_c|$  map. A bulk solvent mask and an anisotropic *B*-factor correction were also used throughout this refinement. The residual Fourier synthesis calculated after final energy minimization and tightly restrained individual *B*-factor refinement of this model presented features in the active site of the docked form of the RNA (chain A) that were consistent with the benzyl group of PheEE (Supplementary Fig. 3). This second model comprises both RNA molecules (92 residues each), two selenomethionyl U1A-RBDs (residues 2–92 and 1–91), ten Mg<sup>2+</sup> ions, one K<sup>+</sup> ion and no water molecules (5,366 non-hydrogen atoms). Ramachandran analysis of the model from this second refinement shows that 78.9% and 20.5% of the protein residues have most favoured and additionally allowed backbone conformations, respectively; there are no residues with a disallowed conformation.

**Mutational and *in vitro* selection analysis.** Reselection has been described previously<sup>21</sup>. Activity assays were performed in triplicate as described<sup>21</sup>.

- Ferré-D'Amaré, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
- Ferré-D'Amaré, A. R. & Doudna, J. A. Use of *cis*- and *trans*-ribozymes to remove 5' and 3' heterogeneities from milligrams of *in vitro* transcribed RNA. *Nucleic Acids Res.* **24**, 977–978 (1996).
- Rupert, P. B. & Ferré-D'Amaré, A. R. Crystallization of the hairpin ribozyme: illustrative protocols. *Methods Mol. Biol.* **252**, 303–311 (2004).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61**, 458–464 (2005).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, 2002).
- Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

**A**re you a 'maverick' scientist? Would you like to be? As young researchers try to make their mark in a time of increasing competition and burgeoning interest in science across the globe, they may be wise to look back at how past scientists overcame scepticism and derision to rewrite the textbooks.

In *Rebels, Mavericks, and Heretics in Biology*, Oren Harman of Bar-Ilan University in Israel and Michael Dietrich of Dartmouth College, New Hampshire, bring together 19 essays on iconoclastic biologists that might prove instructive (see *Nature* **454**, 28–29; 2008). Examples include Cyril Darlington, whose investigations into how chromosomes behave during meiosis challenged accepted concepts in cell biology, and Carl Woese, who dared to propose new kingdoms of microbial life.

Categorizing 'types' of rebel is not easy. As Harman and Dietrich note in the 11 July issue of *The Chronicle of Higher Education*, conventional biologists are conventional in the same way, but each maverick has his or her own way of rebelling. Yet there are lessons here for any young researcher who aspires to a Nobel prize. Then, as now, iconoclastic discoveries came at the intersections of disciplines — microbiology and evolutionary biology in Woese's case. And biologists were sometimes cutting-edge even as they emulated views or conceptual frameworks of the past. The twentieth-century Darlington, for example, took a theory-based approach more akin to nineteenth-century biology.

The book does not address more recent mavericks such as Craig Venter, who left the government sector for the private sector to pursue an unconventional genome-sequencing method. Perhaps this demonstrates a new kind of maverick: one driven not just by new ideas, but by new technology and funding models.

Regardless, the maverick who makes his or her mark will need to be a stubborn risk-taker, with a knack for piecing together under-appreciated clues — and a willingness to challenge views that may underpin an eminent scientist's reputation. Are you a maverick? Would you like to be?

**Gene Russo is editor of *Naturejobs*.**

## CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**

Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)  
**Southwest UK/RoW:** Nils Moeller (4953)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)  
**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)  
**North Germany/The Netherlands:**  
Reya Silao (4970)  
**South Germany/Austria:**  
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:** Dennis Chu

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718

Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**  
Vikas Chawla (+91 1242881057)  
e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe (+81 3 3267 8765)  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)  
**Business Development Manager, Greater China/Singapore:**  
Gloria To (+852 2811 7191)  
e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# MOVERS

**Ernst-Ludwig Winnacker, secretary-general, Human Frontier Science Program Organization, Strasbourg, France**



**2007-09:** Secretary-general, European Research Council, Brussels, Belgium

**1998-2006:** President, DFG, Bonn, Germany

**1984-97:** Director, Laboratory of Molecular Biology, Gene Center, University of Munich, Munich, Germany

At the age of 14, Ernst-Ludwig Winnacker was an accomplished pianist and aspiring conductor. He was also a budding scientist who did experiments on fruitflies as a hobby. Pursuing wide-ranging interests, and building the research infrastructure to expand them continent-wide, has been the crux of his career. Next year, he will bring his boundless curiosity to the Human Frontier Science Program Organization (HFSPO) as its secretary-general.

Pursuing a PhD in chemistry at the Swiss Federal Institute of Technology in Zurich, Winnacker took part in a sizable effort to chemically synthesize vitamin B<sub>12</sub>. Following that, he took a postdoc at the University of California, Berkeley, to work with Horace Barker, who discovered the active form of vitamin B<sub>12</sub>. Winnacker set out to isolate enzymes involved in B<sub>12</sub> synthesis, but instead he became intrigued by the use of recombinant DNA and associated techniques to synthesize and manipulate DNA. "Berkeley was my transition from a pure chemist to a biologist," says Winnacker.

While at the Karolinska Institute in Stockholm, Sweden, and then the Institute of Biochemistry at Ludwig Maximilian University of Munich, Germany, Winnacker studied the replication of DNA viruses, an esoteric topic at the time. He convinced the German government to fund the University of Munich's Gene Center, but then started receiving offers for high-powered science-policy positions.

Winnacker was vice-president of the DFG, Germany's main grant-funding agency, when the country was reunified in 1990 and he tried to smooth the process for former East German institutes. He later assumed the DFG presidency and focused on expanding research opportunities. As secretary-general of the new European Research Council, he has had the chance to do this throughout Europe. Investigator grants were recently awarded to several countries (see *Nature* 453, 975-976; 2008).

In all his policy-making positions, Winnacker has been an ardent defender of, and creative force behind, research programmes designed to support young people, women in science and international collaborations, says Dieter Imboden, president of the Research Council of the Swiss National Science Foundation. At the HFSPO, Winnacker plans to strengthen its focus on interdisciplinary approaches to solving problems. But he believes in funding risky, novel collaborations. "We want to fund systemic approaches to science — this is the future," he says. ■  
**Virginia Gewin**

## SCIENTISTS & SOCIETIES

### Recruiting by rail

As Andreas Klauke rolled through a 15,000-kilometre, 57-city tour of India on a train laden with scientific experiments and displays, he was greeted by cows on the platform, elephants in the street and a total of 2.2 million visitors.

For Klauke, project manager of the Max Planck Society's Science Express, the seven-month trip served as both educational outreach and a massive recruitment effort. The German society launched the project to foster collaborations and recruit students and researchers from India's billion-plus population — especially those who might do their PhD research with the Max Planck Society.

When the Indian government and the Munich-based society first discussed joint scientific outreach, they were unsure how to get to people in far-flung places. India has few museums, and moving exhibits from city to city would have been time-consuming and cumbersome. Then one Indian official recalled a travelling train exhibition that celebrated the country's independence in 1947. Someone jokingly suggested the same for the society's exhibition.

It proved a wise move. Once set up, the 400-metre-long train's 12 themed exhibitions — ranging from the origins of the Universe to three-dimensional

protein animations — could stay intact for the entire trip. It reached small villages as well as large cities. Besides the occasional large mammal blocking the rails, the Science Express was sometimes slowed by the unexpected turnouts. "When 10,000-20,000 people showed up at some stations, the platform wasn't always big enough to handle the crowds," Klauke says.

A science infrastructure gap presents the biggest obstacles to Indian and German scientists working more closely together, Klauke says. He visited some research institutions and universities that had state-of-the-art molecular-biology labs next to labs with outdated equipment. "On one side, they make experiments by hand," he says. "On the other side, they have instruments to go deep into molecular structure."

Hoping to expand the reach of the programme, the Indian government will support another leg of the Science Express's trip in about three months, after some exhibits have been changed or updated. This time, Indian scientists will run the experiments and conduct outreach. And the journey will reach about 50 cities, mostly in the interior, an area not widely covered during the first tour. ■  
**Paul Smaglik**

#### POSTDOC JOURNAL

### Taking steps

One step forward, one step back — let's do the random walk. I'm the DJ of my evolution computer simulations. I make the rules, I set the parameters, and the individuals in my simulated landscapes dance. The number of individuals of each species jumps up and down as time passes, performing a dance that's technically known as a random walk. Ultimately this mix of procreation and chance leads to the ascendance of one species, which I record before moving on to the next parameter set. And so it has been for the past month. Simulation, watch them dance, result. Adjust parameters, watch them dance...

The success of my simulated individuals is not completely random. Yes, chance is a large component, but individuals differ ever so slightly and, given time, these differences become apparent. They underlie Darwin's insights into the wonderful diversity of life and, I suspect, the long-term trend in my stochastic daily life. For every ascendant idea that I have there are countless flops, just as each working computer program hides a trail of error-strewn code. Academic jobs appear and I apply with no result. How many applications does it take to make a career step? A quick survey of my colleagues suggests more than five. I now have six under my belt. Last week I achieved the next step in my career path. Just one step and I'm becoming a lecturer. Now let me dance. ■

**Jon Yearsley is a senior postdoc in evolutionary genetics at the University of Lausanne in Switzerland.**



# Ignorantia juris

Beware the long arm of the law.

**Gareth Owens**

There was always a remote possibility of a time traveller turning up from the moment the Large Hadron Collider went live, but judging by the looks on their faces, I got the feeling that they weren't really expecting one.

Scientists are like that. They'll tell you they think something is possible, like time travel or life on Mars, just to get you excited about funding them, but they don't really believe it themselves because they don't believe in magic. However, much as they'd have liked to, they were never able to find the bit in the details of the Universe that would render spooky old time travel impossible.

I arrived in a puff of something that looked like smoke and felt like semolina, standing with my briefcase held before me like a shield. For me, I neither need to believe nor disbelieve — I'm a lawyer — and on this trip I was a paying passenger of the Temporal Railroad Company of 2177.

If I'd landed on target I'd be standing on the minus side of the Compact Muon Solenoid, under the main shaft. I looked around, still feeling queasy. The green paint of the aluminium inspection platform seemed impossibly bright and the smell of concrete and ozone was exactly like the mock-up in the time station. I blinked a couple of times, and the relief forced the tension to let go of my shoulders. I had made it.

The LHC was the first machine to create subatomic temporal distortions, no good for time travel in themselves, but they formed the zero datum point from which a time machine existed, and so the later technologies based on low-dimensional temporal geometry were able theoretically to exploit this as the end of a railway line. I say 'theoretically' because in 2050, my year of travel, nobody had ever gone back that far. When I told the company my intended destination they flatly refused, at first.

The LHC, they explained, isn't so much the last station on the line, as the buffers at the end of it. It was all Geek to me, and they could throw as much of it at me as they liked. I had a place and a time and a very special client, and they were going to send

me there, and my company was not one to be denied lightly.

My watch had already been set to local time. I had a little bit of running to do, then three minutes of hiding, before I could walk out on the back of a tour group. My briefcase contained everything I needed to blend into the background. I was amazed at the calm faces of the staff I saw as I moved



around the place. They had built this completely amazing machine, designed to throw energies around on a scale previously never envisaged, and they had no expectation of anything amazing happening, like a lawyer from the future falling out of a wormhole.

I had planned my exfiltration based on camera timings and security logs from the online archive. I was particularly keen not to get caught. I hadn't come back 40 years to spend the next 40 rendered to an existence-denied prison somewhere extraordinary.

Aside from a few technological items that were meant to ease my escape from Switzerland, my briefcase was also full of period traveller's cheques and temporally valid ID, so I left CERN behind relatively easily and made my way to Amsterdam.

I knew where to find her because of the arrest records. I got there about an hour before she was due to be raided for drugs.

Her name was Barbi Bodega, and she was worth all the risks. She was lead singer of Franken Ken. I leaned on the bell until she emerged into the pale summer sunlight. She stood in the doorway, a cigarette drooping from her famous pout and a child held to her hip. She was just the way she was in the photos: her hair bleached and dyed so many times that the texture was like carpet, her skin corpse white. She smelled of vomit and English gin.

"Consider yourself served," I said, giving her the papers. She squinted at the summons.

"I'm being sued?" her voice, contralto and wrecked by alcohol, broke with the surprise. "What for? Who by?"

I reached out and tickled the baby's chin.

"Barbi Bodega 2050," I said.

"What?" She really didn't have a clue. Nothing like the woman I remembered.

"Barbi Bodega in 2050 is 62 years of age," I said. "She is a radically different person from you, and she has inherited the body that you created for her, and all the medical problems that entails. To be honest I can't see her surviving another summer like the last one. And she wants what you owe her, another 20 years of life. So she's suing you for reckless endangerment and is applying for sequestration of all the profits from your last two albums."

"You're having a laugh ain'tcha? Lawyer from the future? No such thing."

"Oh I admit, this will be the first time your legal system has come across one, but I have been temporal specialist with Boylett, Wisty and Grole for the past 20 years, and most of the legislation I need is already in place. The Temporal Bar Association calls it accidental case-law. Trust me on this, I'm going to win, and my Barbi Bodega is going to get what the people who care about her really want for her ... more life."

"But you can't take all of my money. How am I going to live? I've got a kid now, how am I supposed to look after him?"

I turned and walked away.

"Don't worry about him," I said. "He grows up to be a lawyer."

**Gareth Owens, trained in Babylonian wizardry, lives under a huge yew tree with a one-eyed cat called Freddy, where he drinks too much coffee and makes stuff up.**

JACEY